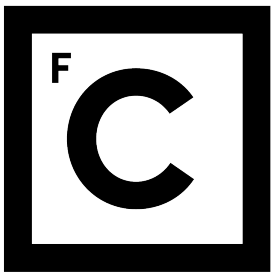
UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS



Ciências
ULisboa

Stochastic methods for the characterization and simulation of wind energy production

Doutoramento em Física

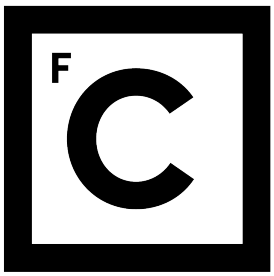
Teresa Scholz

Tese orientada por:
Dr. Vitor Lopes, Dr. Pedro Lind, Prof.Dr. Margarida Telo da Gama

Documento especialmente elaborado para a obtenção do grau de doutor

2016

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS



Ciências
ULisboa

Stochastic methods for the characterization and simulation of wind energy production

Doutoramento em Física

Teresa Scholz

Tese orientada por:
Dr. Vitor Lopes, Dr. Pedro Lind, Prof.Dr. Margarida Telo da Gama

Júri:
Presidente:

- José Manuel de Nunes Vicente e Rebordão

Vogais:

- Matthias Wächter
- Philipp Maass
- Pedro Gonçalves Lind
- João Pedro da Silva de Brito Boto
- Nuno Miguel Azevedo Machado de Araújo

Documento especialmente elaborado para a obtenção do grau de doutor

Fundação para a Ciência e Tecnologia (SFRH/BD/86934/2012, PTDC/SEN-ENR/114178/2009, FCOMP-01-0124-FEDER-016080, DRI/DAAD/1208/2013, UID/FIS/00618/2013), German Academic Exchange Service (DRI/DAAD/1208/2013), IPID4all programme supported by the German Academic Exchange Service with funds from the Federal Ministry of Education and Research

2016

Contents

Acknowledgements

I would like to gratefully acknowledge

- the Fundação para a Ciência e Tecnologia (SFRH/BD/86934/2012, PTDC/SEN-ENR/114178/2009, FCOMP-01-0124-FEDER-016080, DRI/DAAD/1208/2013, UID/FIS/00618/2013),
- the German Academic Exchange Service (DRI/DAAD/1208/2013) and
- the IPID4all programme supported by the German Academic Exchange Service with funds from the Federal Ministry of Education and Research

for financial support and GENERG, SA for providing me with the original wind data.

The completion of this PhD thesis was only possible with the support of several people. I would like to express my sincere gratitude to all of them:

Vitor Lopes for his patient guidance, encouragement and advice throughout my whole time as his student- from the "nursery" during my Master's degree to this PhD thesis. The joy and enthusiasm he has for his research were contagious and motivational for me. I have been lucky to have had a supervisor who cared so much about my work while at the same time giving me the freedom to do things in my own "black-and-white" way. During this time he taught me a lot about data analysis and modeling, programming, paper writing and real-world-stuff, such as the stratification of tea or that the only tools you need to raise a cat are a bag full of friskies and starvation.

Pedro Lind, who has been a true "Doktorvater" to me, an inexhaustible source of ideas and an inspiring teacher. His apparently endless energy paired with his genuine friendliness and humor make working with him a pleasure. Throughout this whole PhD he was generous with both praise and constructive criticism and he always made time to help me answer my questions regardless of his busy schedule. In addition to his encouragement and guidance that helped me to grow as a researcher I would also like to thank him and the whole *LaLi*-clan for all the help, support and fun (!) in life outside of faculty, be it day trips to the countryside, chatting about marathon training or indulging in German-ness with *Feuerzangenbowle*, *Mohnkuchen* and *Kartoffelsalat*.

Frank Raischel, who has known the answer to every question I have ever dared to ask about stochastic processes, programming or provisioning of Bavarian food in Portugal.

Margarida Telo da Gama for the opportunity to do this PhD thesis, for welcoming me into the CFTC research institute and providing me with all necessary installations and documents.

The research group at the ForWind center for welcoming me there.

André Ferreira for being my tank, support and healer as much in real life as in Sanctuary, Midgård or Tamriel.

My whole family for their constant love and support, above all my parents, to whom I owe everything.

Abstract

The aim of this PhD thesis is to apply tools from stochastic modeling to wind power, speed and direction data, in order to reproduce their empirically observed statistical features. In particular, the wind energy conversion process is modeled as a Langevin process, which allows to describe its dynamics with only two coefficients, namely the drift and the diffusion coefficients. Both coefficients can be directly derived from collected time-series and this so-called Langevin method has proved to be successful in several cases. However, the application to empirical data subjected to measurement noise sources in general and the case of wind turbines in particular poses several challenges and this thesis proposes methods to tackle them.

To apply the Langevin method it is necessary to have data that is both stationary and Markovian, which is typically not the case. Moreover, the available time-series are often short and have missing data points, which affects the estimation of the coefficients. This thesis proposes a new methodology to overcome these issues by modeling the original data with a Markov chain prior to the Langevin analysis. The latter is then performed on data synthesized from the Markov chain model of wind data. Moreover, it is shown that the Langevin method can be applied to low sample rate wind data, namely 10-minute average data.

The method is then extended in two different directions.

First, to tackle non-stationary data sets. Wind data often exhibits daily patterns due to the solar cycle and this thesis proposes a method to consider these daily patterns in the analysis of the time-series. For that, a cyclic Markov model is developed for the data synthesis step and subsequently, for each time of the day, a separate Langevin analysis of the wind energy conversion system is performed.

Second, to resolve the dynamical stochastic process in the case it is spoiled by measurement noise. When working with measurement data a challenge can be posed by the quality of the data in itself. Often measurement devices add noise to the time-series that is different from the intrinsic noise of the underlying stochastic process and can even be time-correlated. This spoiled data, analyzed with the Langevin method leads to distorted drift and diffusion coefficients. This thesis proposes a direct, parameter-free way to extract the Langevin coefficients as well as the parameters of the measurement noise from spoiled data. Put in a more general context, the method allows to disentangle two superposed independent stochastic processes.

Finally, since a characteristic of wind energy that motivates this stochastic modeling framework is the fluctuating nature of wind itself, several issues raise when it comes to reserve commitment or bidding on the liberalized energy market. This thesis proposes a measure to quantify the risk-return-ratio that is associated to wind power production conditioned to a wind park state. The proposed state of the wind park takes into account data from all wind turbines constituting the park and also their correlations at different time lags.

Keywords: Stochastic process, Markov chain, Langevin analysis, Measurement noise, Wind energy

Resumo

Esta tese visa a aplicação do método de Langevin a dados da potência eólica, da velocidade e da direção do vento. Modelar o processo de conversão da energia eólica como um processo de Langevin permite descrever a sua dinâmica, utilizando unicamente dois coeficientes, nomeamente os coeficientes de arraste e de difusã, que representam o comportamento determinístico e estocástico do sistema, respetivamente. Ambos os coeficientes podem ser derivados diretamente das séries temporais medidas. No entanto, a aplicacão a dados medidos, em geral, e a dados medidos em turbinas eólicas, em particular, coloca vários desafios. Esta tese propõe métodos para os resolver.
Para analisar dados com o método de Langevin é necessário que sejam estacionários e Markovianos, o que *a priori* não é dado. Além disso, as séries temporais disponíveis são tipicamente muito curtas e com frequentes intervalos sem dados disponíveis. Ambos estes problemas influenciam o processo de estimativa. Esta tese propõe uma nova metodologia para ultrapassar estes obstáculos, utilizando a modelação dos dados originais com uma cadeia de Markov previamente à análise de Langevin. Esta ultima é realizada com dados sintetizados a partir da cadeia de Markov. Estes dados sintéticos não têm falhas de pontos de dados, são de extensão arbitrária e, por construção, Markovianos e estacionários.
Ao trabalhar com dados medidos, outro desafio é colocado pela própria qualidade dos dados. Frequentemente, os dispositivos de medida adicionam um ruído às séries temporais que é de natureza diferente do ruído intrínseco do processo subjacente estocástico. Analisar estes dados contaminados com o Método de Langevin pode ser difícil, dado que o ruído que está presente nos dados também distorce os coeficientes de arraste e difusão extraídos destes. Esta tese propõe uma forma directa, não paramétrica, de estimar os coeficientes de Langevin, assim como os parâmetros do ruído de medida.
Uma situação frequente quando se lida com medições de dados de vento é a sua baixa taxa de amostragem. Esta tese aplica o método de Langevin aos dados médios de periodos de 10 minutos, que são os dados tipicamente disponíveis, e mostra que mesmo com esta baixa taxa de amostragem, é possível extrair as características da dinâmica do processo de conversão de energia eólica.
Dados de vento exibem muitas vezes padrões diários devido ao ciclo solar e esta tese propõe um método para considerar estes padrões diários na análise das séries temporais. Para isso, desenvolve-se um modelo cíclico de Markov para a sintetização dos dados e subsequentemente, para cada periodo de 10 minutos do dia, é feita uma análise de Langevin do sistema de conversão de energia eólica.
Outra característica da energia eólica é a natureza intermitente da sua fonte (daí a modelação através de processos estocásticos), que levanta várias questões no âmbito das reservas contratuadas no mercado liberalizado da energia. Esta tese propõe uma medida para quantificar o quociente de risco-retorno que está associado à produção de energia eólica, condicionado ao estado do parque eólico. O estado do parque eólico definido toma em conta dados vindos de todas as turbinas eólicas do parque e também as suas correlações temporais.

Palavras-chave: processo estocástico, cadeia de Markov, análise de Langevin, ruído de medida, energia eólica

Part I

Background

Chapter 1

Introduction and overview

The demands for electric energy has increased dramatically over the last century and wind power has become a competitive alternative to conventional energy sources. It is inexpensive, renewable, produces no greenhouse gases or hazardous waste and uses little land [Gasch and Twele, 2012]. Its drawback however is its high variability, which imposes a challenge to incorporating it into a grid. To compensate for this variability, wind energy is usually combined with other power sources, such as water or gas (reserve commitment) and the industry relies on weather forecasts to predict the power output and to act accordingly.

Atmospheric wind is caused by pressure differences across the earth's surface due to uneven heating of the earth by solar radiation. The regions close to the equator absorb more solar radiation than for example the regions around the poles. These spatial variations in heat create pressure variations that cause winds to flow from high to low pressure areas [Burton et al., 2001]. This results in a large-scale global circulation pattern, which is disturbed by variations on continental scales due to geography, i.e. oceans and land masses. On an even smaller scale, the local topography has a huge effect on the nature of the wind. For example, hills and mountains result in regions of increased wind speed whereas sheltered valleys lead to a reduced wind speed. In addition to this, there are thermal effects that can result in local variations such as temperature gradients between altitudes [Manwell et al., 2009] or differential heating between land and sea. Local effects like friction with the earth's surface or variations of temperature also have a strong impact on the wind causing a turbulent flow close to the ground.

This variability of the wind over a wide range of scales exists not only in space but also in time: atmospheric winds exhibit variations on time scales covering several orders of magnitude, ranging from centuries [Böttcher et al., 2006] to seconds or less [Burton et al., 2001]. This co-existence of several scales is one of the main ingredients of what is known as the turbulent character of wind dynamics. Richardson [1922] suggested, that in a turbulent flow, energy is passed down from the larger to the small scales through a hierarchy of decreasing vortices, where it dissipates due to viscous stresses. This concept of an energy cascade was built on by Kolmogorov [1941b,a], who introduced the so-called Kolmogorov scale below which the dissipative effects dominate the dynamics.

Electric energy generated from wind power fluctuates at several different time scales due to this turbulent behavior of the wind and therefore, describing it is an important research topic. This can be achieved using the Navier-Stokes equations, partial differential equations describing any single-phase fluid flow, taking into account pressure, density, viscosity and external forces. Methods from computational fluid dynamics serve to solve these equations numerically for certain initial conditions and subject to proper boundary conditions. However to do that for all scales of interest is computationally very expensive. Moreover, the air flow can be described as chaotic [Burton et al., 2001], i.e. small variations in the initial or boundary conditions can result in large differences in the predictions. In contrast to this chaotic character of the details of wind flow and the fact that they vary from one realization of a flow to the next, its statistical properties seem to be well behaved and well

reproducible [Burton et al., 2001]. Therefore, an alternative approach is to use a probabilistic *ansatz*, which describes the macroscopic properties of the wind flow as a whole instead of following the movement of each single gas molecule. In this context, methods of generalized stochastic processes are a promising tool [Peinke et al., 2014] and both wind velocity and power have been described through Langevin dynamics. Such models assume the wind speed or power at a moment in the future to be the sum of the wind speed or power at the current time and a noise contribution intrinsic to the wind. The coefficients describing the influence of the current time as well as the amplitude of the noise term can be directly estimated from experimental or measurement data [Friedrich et al., 2011]. This *ansatz* proved to be very successful, as described in Chapter 2.
Since the stochasticity of the wind field transfers to the stochastic dynamics of the power output of the wind turbine, the application of the Langevin method to wind energy conversion systems could lead to a better understanding of the stochastic aspects of power production coupled to the wind velocity field.
The application of the Langevin method to wind energy conversion systems is the heart of this thesis. After describing the state of art in Chapter 2, covering the theory of stochastic processes as well as a summary of its applications to the area of wind energy, the main contributions are presented in Part II as follows:

- Chapter 3 introduces a model of a real-world dataset containing measurements of wind power, speed and direction at a wind turbine based on a discrete, finite-state Markov process. The theory of Markov chains is introduced and the modeling process is described in detail. First, a method to partition the state space is presented. Then, an optimization problem is formulated, whose solution yields the model parameters. This is achieved by minimizing a constrained objective function that is a combination of negative log-likelihood functions describing the likelihood of one-, two- and three-step state transitions in the data. The obtained model is then compared to the original data to ensure all statistical properties are captured and two possible applications are suggested: persistence analysis and the validation of additional information in terms of forecast. The paper corresponding to this chapter that introduces on of the fundamental ideas of the methodology presented in this thesis was published in 2012.

- Chapter 4 models and analyses the dynamic behavior of this turbine in terms of the Langevin approach. For this, a synthetic dataset, based on the model described in Chapter 3 is generated. This synthetic data is Markovian by construction, does not have missing data points or periodicities and is of arbitrary length. Thus, this analysis avoids typical issues with real-world data while containing its statistical properties and dynamic behavior. The Langevin approach allows to separate the dynamics of the turbine into a deterministic and a stochastic part, given by the so-called drift and diffusion coefficients at each point in state space. Analysis of the resulting coefficients in the wind-speed plane uncovers specific features of the wind turbine and characterizes its different working regions. Moreover, it is shown, how the power curve, describing the functional dependence between wind speed and power can be derived from both drift and diffusion coefficients.

- Chapter 5 extends the discrete, finite-state Markov model introduced in Chapter 3 to describe an inhomogeneous process, capturing diurnal patterns in wind power production. The time-dependent transition probabilities are modeled using Bernstein polynomials which reduces the problem size. Again, the model parameters are estimated using an optimization problem. Its objective function consists of two parts, one ensuring that the long-term statistical properties of the data are captured and one that incorporates the daily variations. The formulation of the convex optimization problem with all constraints is described in detail. Subsequently, the model is compared to the original data to ensure all statistical properties are captured. Moreover, a time-dependent persistence analysis of power production is presented.

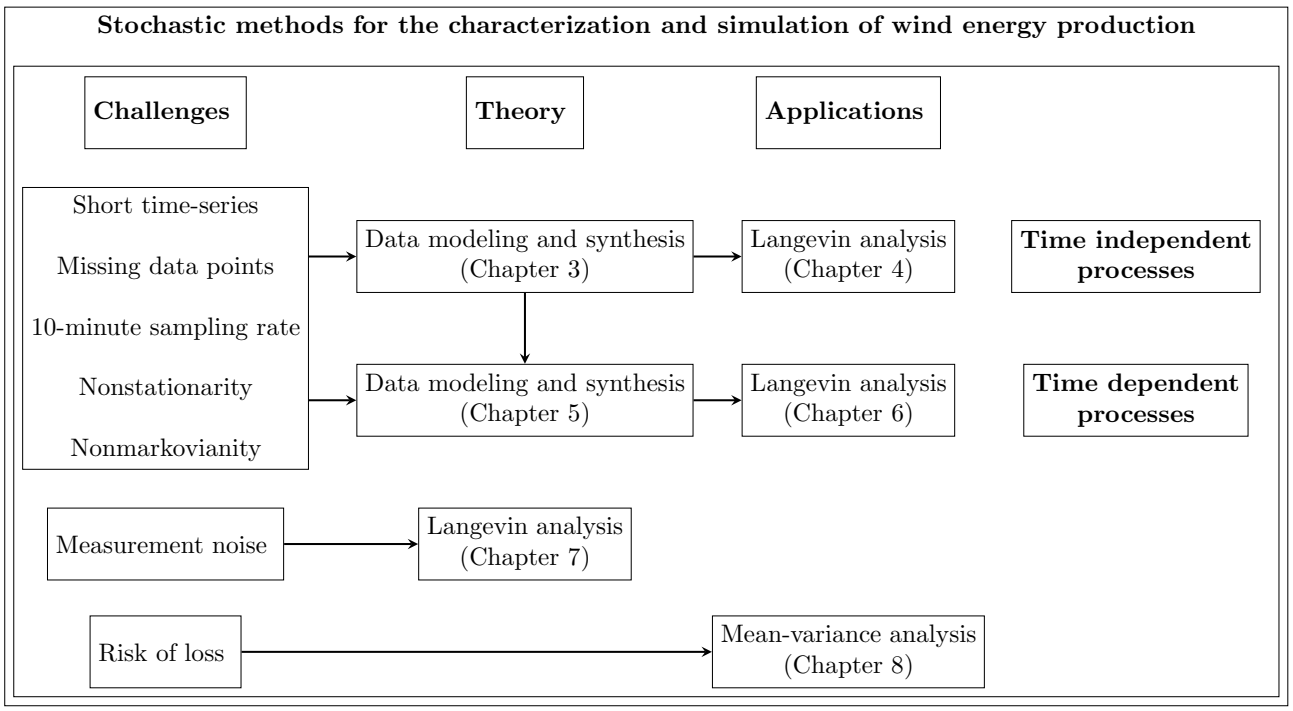


Figure 1.1: Scheme of the contributions presented in this thesis together with the specific challenges they tackle.

- Chapter 6 describes the application of the Langevin method to synthetic data generated with the model introduced in Chapter 5. This has the advantages stated above compared to the direct analysis of the real-world data. The results are then used to investigate changes in the dynamic behavior of the turbine throughout the day, which is demonstrated on an example.
- Chapter 7 addresses a problem that frequently occurs with real world data: Often the collected data is spoiled with measurement noise and directly analyzing this noisy data would lead to distorted drift and diffusion coefficients and thus a distorted description of the system's dynamics. The paper describes a parameter-free algorithm that can separate the unspoiled time-series from this measurement noise, given that the latter can be modeled as an Ornstein-Uhlenbeck process. The method can also be seen as a way to disentangle two stochastic signals. The theory behind the methodology is briefly summarized and its implementation described in detail. The algorithm's functionality is demonstrated on a synthetic dataset.
- Chapter 8 explores the area of energy markets and proposes a modified mean-variance, also called risk-return, approach to power production. Here, a wind park "state" is suggested, based on the velocity field of all wind turbines in a park and at several time lags. It is then used to formulate a risk-return measure conditioned to this state.

Section 9 presents the overall conclusions and discusses some open questions for future work.
The scheme in figure 1.1 presents the contributions of this thesis and their interdependencies in the context of the challenges they tackle.

Chapter 2

State of the art

2.1 An historical example: Brownian motion

London, June, 1827. Robert Brown is in his botanical collection at Soho square that he had taken over from his mentor, Sir Joseph Banks after his death in 1820. His research in the last years had led him to the decision to study the structure of the Pollen and to inquire into its mode of action on the Pistillum, the female part of the flower, since he doubted the observations of his colleagues on the mode of action of the pollen in the process of impregnation. To begin his studies, he had chosen *Clarckia pulchella*, a plant whose pollen were filled with particles or granules of unusually large size and of a figure between cylindrical and oblong. He carefully prepared a sample of the pollen in water and slid it under his brass-built microscope. Briefly he was tempted to use the new lenses he had just obtained, they had much higher power than his usual lens. But to give greater consistency to his statement and also to bring the subject as much as possible within the reach of general observation, he decided to employ the simple lens throughout his inquiries.
As he examined the form of the particles immersed in water, he observed many of them very evidently in motion, a motion not only of a change of place in the fluid but also not infrequently of a change of form in the particle itself. His curiosity was piqued. He prepared more samples and after frequently repeated observation of the movement, he concluded that it did not arise from currents in the fluid or its gradual evaporation but belonged to the particle itself. Could this be true? Was he observing the essence of life?
He extended his studies to other plants of the same family, then numerous other species having particles that varied in form from oblong to spherical. He found motion in the particles of the pollen of all living plants he examined and decided to inquire whether this property continued after the death of the plant and for what length of time it was retained. So he went to his herbarium and prepared samples from specimens of several plants, some of which had been dried and preserved for twenty years and others not less than a century. As he eyed through his microscope, to his big astonishment, he observed the particles in evident motion. Brown then had the suspicion the movement was associated with organic matter so he moved this experiments on to organic bodies, like pit-coal, fossil wood or other substances of vegetable origin. The movement was clearly visible. He went on with his experiments and noted, that his suspicion was wrong and the motion was not limited to organic bodies or even their bodies which he proved by analyzing window glass, samples of metals, rocks of all ages and even a fragment of the Sphinx and meteorites from various localities. Brown concluded, that any solid mineral would reveal the phenomenon once it was reduced to a sufficiently fine powder to be temporarily suspended in water.
Robert Brown was not the first one to observe the motion and following the publication of his observations in 1827 [Brown, 1828] and 1829 [Brown, 1829] many more experiments were conducted. However, only almost 70 years later, Einstein [1905] gave the explanation for the motion. In his 1905 paper he stated that it was a result of the pollen being moved by collisions with water molecules. Independently, Smoluchowski [1906] published the same results a year after. Their

theory of Brownian motion was verified experimentally by Perrin [1913] in 1908 and thus it was confirmed that atoms and molecules actually exist, which so far had only been speculated. But Einstein's paper was ground-breaking in another sense as well: With the probabilistic description of the fluid's molecules' effect on the pollen grain it was the beginning of stochastic modeling of natural phenomena.
Many of the concepts employed in this thesis can already be found in Einstein's paper. Assuming that the "push" of a particle at a given time is independent of any previous motion he uses the so-called Markov property and derives a special form of the Chapman-Kolmogorov equation. From there, using an approximation that is essentially the Kramers-Moyal expansion, he arrives at a diffusion equation, which is a special case of the Fokker-Planck equation [Gardiner, 2009]. Years after Einstein's first description, Langevin presented a different and according to him "infinitely more simple" method to approach Brownian motion and in the course of his paper gave the first example of a stochastic differential equation, thereafter known as the Langevin equation [Langevin, 1908]. All the mentioned approaches that were first presented with the investigation of Brownian motion are essential to the subject of stochastic processes, which are an important tool for the description of complex systems. Assuming a basic knowledge of probability theory [Feller, 1968, 1971], this chapter will formally introduce the theoretical concepts that are fundamental to this thesis. It is based on the well-known books on the subject of stochastic processes by Gardiner [2009] and Risken [1984].

2.2 Stochastic process

The previous section introduced Brownian motion as the first example of a stochastic process, which is the evolution of a random variable[a] $X(t)$ representing the state of some system over time, indexed by t. In general, a stochastic process can be multidimensional, i.e. $X(t) \in \mathbb{R}^N$, however, this introduction focuses on one dimensional stochastic processes $X(t) \in \mathbb{R}$. The multidimensional case can be described in analogue, the only changes that arise with the transition to N dimensions are described in Section 2.6. Stochastic processes, as dealt with in this thesis, are considered to be time-continuous, but their state $x_i = X(t_i)$ can only be observed at discrete times t_i, $i \in [0,...,T]$ for some $T \in \mathbb{N}$. Using these observations or measurements, a complete description of the stochastic process can be given by a set of joint probability densities $f(x_0,t_0;x_1,t_1;\ldots;x_T,t_T)$. Based on the joint probability densities, conditional probability densities are defined as

$$p(x_{i+1},t_{i+1}|x_i,t_i;\ldots;x_0,t_0) = \frac{f(x_{i+1},t_{i+1};x_i,t_i;\ldots;x_0,t_0)}{f(x_i,t_i;\ldots;x_0,t_0)} \tag{2.1}$$

and can be viewed as a probabilistic prediction for the future value $X(t_{i+1})$ of the process knowing its values $X(t_i),\ldots,X(t_0)$ in the past. This interpretation of the conditional probability as the "memory" of the stochastic process leads to the definition of a special class of stochastic processes, the so-called Markov processes.

2.3 Markov process

A Markov process is a memory-less stochastic process, i.e. the probability of any future state of the system only depends on its current state, not on the past. In terms of conditional probabilities this

[a] A remark on notation: In this thesis, the capital letter is used to describe the stochastic process, whereas the small letter is used to refer to a specific value, an observation of the stochastic process, in state space. Therefore, the values x_i are in general not in order ($i < j \not\Rightarrow x_i < x_j$) and they can assume the same values ($x_i = x_j \not\Rightarrow i = j$). It is just for simplicity that the values are indexed with the same index as the time of their observation. Thus, $X(t_i)$ is variable that can assume any value in the state space, x_i is a certain value in state space that in the trajectory under observation is assumed by X at time t_i.

Markov property can be formulated as

$$p(x_{i+1},t_{i+1}|x_i,t_i;\ldots;x_0,t_0) = p(x_{i+1},t_{i+1}|x_i,t_i). \tag{2.2}$$

Applying the Markov property to equation (2.1) leads to

$$f(x_{i+1},t_{i+1};x_i,t_i;\ldots;x_0,t_0) = f(x_0,t_0)\prod_{j=0}^{i} p(x_{j+1},t_{j+1}|x_j,t_j), \tag{2.3}$$

which shows that a Markov process can be completely described using only the conditional probabilities $p(x_{i+1},t_{i+1}|x_i,t_i)$, $i \in [1,...,T]$ and its initial state $x_0 = X(t_0)$. The law of total probability together with the Markov property yields the central dynamic equation to all Markov processes, the so-called Chapman-Kolmogorov equation

$$p(x_{i+1},t_{i+1}|x_{i-1},t_{i-1}) = \int_{-\infty}^{\infty} p(x_{i+1},t_{i+1}|x_i,t_i)p(x_i,t_i|x_{i-1},t_{i-1})dx_i, \tag{2.4}$$

which states that the probability of the system transitioning to a state x_{i+1} from x_{i-1} is given as the integral over all possible states the system can assume in between the corresponding time-steps t_i.

2.4 Discrete Markov process

In the special case of a discrete, finite, state space $\mathscr{S} = \{s_0,s_1,...,s_n\}$, $n \in \mathbb{N}$, the Markov property reads

$$p(s_j,t_{k+1} \mid s_i,t_k;s_{l_{k-1}},t_{k-1};\ldots;s_{l_0},t_0) = p(s_j,t_{k+1} \mid s_i,t_k). \tag{2.5}$$

All transition probabilities $p_{ij}(t_k) = p(s_j,t_{k+1} \mid s_i,t_k)$ can be described by the so-called transition matrix $\mathbf{P}^{(t_k)}$, a $n \times n$ matrix with entries $P_{ij}^{(t_k)} = p_{ij}(t_k)$ for all $i,j \in \{0,\ldots,n\}$. Each transition probability fulfills the properties $p_{ij}(t_k) \geq 0$ and $\sum_j p_{ij}(t_k) = 1$. The Markov process is completely described by these transition matrices and the initial state value $X(t_0)$ due to the Chapman-Kolmogorov equation, that for a discrete state space reads:

$$p(s_j,t_{k+1} \mid s_i,t_{k-1}) = \sum_l p_{il}(t_{k-1})p_{lj}(t_k), \quad \forall i,j \in \{0,\ldots,n\}. \tag{2.6}$$

If the transition probabilities repeat themselves after a certain period T, i.e. if T is the smallest number, such that $p_{ij}(mT+r) = p_{ij}(r)$ for integer m and $0 \leq r < T$ [Platis et al., 1998], the Markov process is called cyclic and can be described by its initial distribution and T transition matrices $\mathbf{P}^{(r)}$, $r = 0,...,T-1$. If the transition probabilities are constant and thus the process can be completely described by one transition matrix and its initial state value, it is called a (time-homogeneous) Markov chain [Kemeny and Snell, 1976].
An interesting property of any Markov process is its long-term behavior, in particular how much time it spends in each state as t approaches infinity. For a discrete finite state space this long-term behavior is described by the limiting distribution $\pi = (\pi_0,\ldots,\pi_n)$, where $\pi_j = \lim_{t\to\infty} p(s_j,t)$. If the Markov chain is aperiodic and it is possible to reach every state from any other state, this distribution is stationary and it can be computed as $\pi = \lim_{t\to\infty} \mathbf{P}^t$. It is independent from the initial distribution of the Markov chain and satisfies the balance equation $\pi = \pi\mathbf{P}$. By the Perron-Frobenius theorem it can be computed as the normalized eigenvector corresponding to the unit eigenvalue of the transition matrix [Pillai et al., 2005].
In the case of a cyclic time-inhomogeneous Markov process, for each $0 < r < T$ a stationary distribution exists, which can be interpreted as the limiting distribution of the Markov process considering only the data points at time steps that are a multiple of r. The stationary distribution π_r is given by

$$\pi_r = \pi^* \cdot \mathbf{P}^{(0)} \cdot \ldots \cdot \mathbf{P}^{(r-1)}, \tag{2.7}$$

where $\pi^* = \lim_{t\to\infty} \mathbf{P}^t$ and $\mathbf{P} = \mathbf{P}^{(0)} \cdot \ldots \cdot \mathbf{P}^{(T-1)}$. Details of the derivation can be found in Chapter 5.

2.5 Kramers-Moyal coefficients and the Fokker-Planck equation

A stochastic process can evolve in many (often infinitely many) ways, even when starting from the same initial state. However, even though the realization of the stochastic process at a time t is random, the probability of each possible value is fixed. Therefore, the time evolution of the stochastic process can (and should) be described in a probabilistic way, i.e. by the evolution of its distribution function. For the case of Markov processes it is given by the so called Fokker-Planck equation, which will be introduced in this section. The derivation follows the approach presented by Risken [1984], who uses the conditional moments of the Markov process to define the Kramers-Moyal expansion of the distribution function and then, employing the theorem of Pawula, proves that it reduces to the Fokker-Planck equation. A different approach can be found in Gardiner [2009], deriving the Fokker-Planck equation from a differential version of the Chapman-Kolmogorov equation and thus directly deriving it from the Markov property.
The Kramers-Moyal expansion is an expansion of the distribution function that makes use of the moments of the stochastic process, defined as

$$M_n(x,t,\tau) = \langle [X'(t+\tau) - X'(t)]^n | X'(t) = x \rangle = \int_{-\infty}^{\infty} (x'-x)^n p(x',t+\tau|x,t) dx', \tag{2.8}$$

where τ is a multiple of the discretization time-step Δt and $\langle \cdot | X(t) = x \rangle$ symbolizes a conditional averaging over the entire measurement period, where only measurements with $X(t) = x$ are taken into account.
The differential of the probability distribution of the stochastic process is given as

$$\frac{\partial f(x,t)}{\partial t} = \lim_{\tau \to 0} \frac{f(x,t+\tau) - f(x,t)}{\tau}. \tag{2.9}$$

For the distribution at time $t+\tau$ the law of total probability yields

$$\begin{aligned} f(x,t+\tau) &= \int_{-\infty}^{\infty} f(x,t+\tau;x',t) dx' \\ &= \int_{-\infty}^{\infty} p(x,t+\tau|x',t) f(x',t) dx', \end{aligned} \tag{2.10}$$

where the transition probability $p(x,t+\tau|x',t)$ can be described in terms of the δ-function as

$$p(x,t+\tau|x',t) = \int_{-\infty}^{\infty} \delta(y-x) p(y,t+\tau|x',t) dy. \tag{2.11}$$

Using the Taylor expansion of the δ function

$$\begin{aligned} \delta(y-x) &= \delta(x'-x+y-x') \\ &= \sum_{n=0}^{\infty} \frac{(y-x')^n}{n!} \left(\frac{\partial}{\partial x'} \right)^n \delta(x'-x) \\ &= \sum_{n=0}^{\infty} \frac{(y-x')^n}{n!} \left(-\frac{\partial}{\partial x} \right)^n \delta(x'-x), \end{aligned} \tag{2.12}$$

Equation (2.11) can be reformulated as

$$\begin{aligned} p(x,t+\tau|x',t) &= \sum_{n=0}^{\infty} \frac{1}{n!} (-\frac{\partial}{\partial x})^n \left[\int_{-\infty}^{\infty} (y-x')^n p(y,t+\tau|x',t) dy \right] \delta(x'-x) \\ &= \left(1 + \sum_{n=1}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial x} \right)^n M_n(x',t,\tau) \right) \delta(x'-x) \\ &= \left(1 + \sum_{n=1}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial x} \right)^n M_n(x,t,\tau) \right) \delta(x-x'). \end{aligned} \tag{2.13}$$

Inserting Equations (2.13) and (2.11) into Equation (2.9) leads to

$$\begin{aligned}\frac{\partial f(x,t)}{\partial t} &= \lim_{\tau\to 0}\frac{1}{\tau}\left[\int_{-\infty}^{\infty}\left(1+\sum_{n=1}^{\infty}\frac{1}{n!}\left(-\frac{\partial}{\partial x}\right)^n M_n(x,t,\tau)\right)\delta(x-x')f(x',t)dx' - f(x,t)\right]\\ &= \lim_{\tau\to 0}\frac{1}{\tau}\frac{1}{n!}\sum_{n=1}^{\infty}\left(-\frac{\partial}{\partial x}\right)^n\int_{-\infty}^{\infty}\delta(x-x')M_n(x,t,\tau)f(x',t)dx'\\ &= \lim_{\tau\to 0}\frac{1}{\tau}\frac{1}{n!}\sum_{n=1}^{\infty}\left(-\frac{\partial}{\partial x}\right)^n M_n(x,t,\tau)f(x,t)\\ &= \sum_{n=1}^{\infty}\left(-\frac{\partial}{\partial x}\right)^n\left[\frac{1}{n!}\lim_{\tau\to 0}\frac{1}{\tau}M_n(x,t,\tau)\right]f(x,t).\end{aligned} \tag{2.14}$$

Applying the definition of the Kramers-Moyal coefficients

$$D^{(n)}(x,t) = \frac{1}{n!}\lim_{\tau\to 0}\frac{1}{\tau}M_n(x,t,\tau) \tag{2.15}$$

yields the Kramers-Moyal expansion

$$\frac{\partial f(x,t)}{\partial t} = \sum_{n=1}^{\infty}\left(-\frac{\partial}{\partial x}\right)^n D^{(n)}(x,t)f(x,t). \tag{2.16}$$

This expansion has infinitely many terms, but it can be shown that for many practical applications the coefficients vanish for $n \geq 3$. This is due to the theorem of Pawula, which states that the expansion either stops after the first or second term or it requires an infinite number of terms. To prove it, insert

$$g_1(x) = (x-x')^n, \tag{2.17a}$$
$$g_2(x) = (x-x')^{n+m}, \tag{2.17b}$$
$$p(x) = p(x,t+\tau|x',t') \tag{2.17c}$$

into the generalized Schwartz inequality

$$\left(\int_{-\infty}^{\infty}g_1(x)g_2(x)p(x)dx\right)^2 \leq \int_{-\infty}^{\infty}g_1^2(x)p(x)dx\int g_2^2(x)p(x)dx, \tag{2.18}$$

which yields

$$\begin{aligned}\left(\int_{-\infty}^{\infty}(x-x')^n(x-x')^{n+m}p(x,t+\tau|x',t')dx\right)^2 \leq &\int_{-\infty}^{\infty}(x-x')^{2n}p(x,t+\tau|x',t')dx\\ &\int_{-\infty}^{\infty}(x-x')^{2n+2m}p(x,t+\tau|x',t')dx,\end{aligned} \tag{2.19}$$

that, rewritten in terms of the moments reads $M_{2n+m}^2 \leq M_{2n}M_{2n+2m}$ for all positive integers n,m. Dividing by τ and considering the limit of $\tau\to 0$ allows to apply the definition of the Kramers-Moyal coefficients given by Equation (2.15), yielding

$$\left((2n+m)!D^{(2n+m)}\right)^2 \leq (2n)!(2n+2m)!D^{(2n)}D^{(2n+2m)} \tag{2.20}$$

for the expansion coefficients with $n,m \geq 1$.
From this it can be seen that if $D^{(2r)} = 0$ for any $r \geq 1$, all coefficients $D^{(n)}(x,t)$ with $n \geq 3$ must vanish. The resulting truncated expansion

$$\frac{\partial f(x,t)}{\partial t} = \left(-\frac{\partial}{\partial x}D^{(1)}(x,t) + \frac{\partial^2}{\partial x^2}D^{(2)}(x,t)\right)f(x,t), \tag{2.21}$$

is known as the Fokker-Planck equation and the first and second Kramers-Moyal coefficients as the drift and diffusion coefficient, respectively.

2.6 The Langevin equation

The first stochastic equation, i.e. a differential equation with a random term, was formulated by Langevin in his study of Brownian motion. He assumed, that two forces act on the Brownian particle: a viscous drag and a fluctuating force which represents the impacts of the molecules of the liquid. This description of Brownian motion as the sum of a deterministic part and stochastic fluctuations can be generalized into the nonlinear Langevin equation

$$\frac{dX}{dt} = h(X,t) + g(X,t)\Gamma(t), \tag{2.22}$$

where Γ is a Gaussian distributed, uncorrelated noise with the properties

$$\langle \Gamma(t) \rangle = 0, \tag{2.23a}$$
$$\langle \Gamma(t)\Gamma(t') \rangle = \delta(t-t'). \tag{2.23b}$$

Equations (2.23) provide an intuitive description of a stochastic force: it ensures that the fluctuations average out in time and that for different times t and t' the corresponding values of the stochastic force are statistically independent, i.e. that there is no time-correlation between values of the stochastic force.

The coefficients of the Langevin equation are given by the first two terms of the Kramers-Moyal expansion

$$h(X,t) = D^{(1)}(x,t), \tag{2.24a}$$
$$(g(X,t))^2 = D^{(2)}(x,t), \tag{2.24b}$$

which connects the Langevin with the Fokker-Planck equation, both describing the stochastic process.

These relations are derived in the following. The definition of the drift coefficient and application of the Itô-calculus [Milan, 2014] yields

$$\begin{aligned} D^{(1)}(x_0,t) &= \lim_{\tau\to 0} \frac{1}{\tau} \langle X(t+\tau) - X(t) | X(t) = x_0 \rangle \\ &= \lim_{\tau\to 0} \frac{1}{\tau} \langle \int_t^{t+\tau} \left[h(X,t') + g(X,t')\Gamma(t') \right] dt' | X(t) = x_0 \rangle \\ &= \lim_{\tau\to 0} \frac{1}{\tau} \langle h(X,t) \int_t^{t+\tau} dt' + g(X,t) \int_t^{t+\tau} \Gamma(t') dt' | X(t) = x_0 \rangle \\ &= h(x_0,t) + g(x_0,t) \lim_{\tau\to 0} \frac{1}{\tau} \int_t^{t+\tau} \langle \Gamma(t') | X(t) = x_0 \rangle \, dt' \\ &= h(x_0,t), \end{aligned} \tag{2.25}$$

where the relations $\int_t^{t+\tau} h(x,t')dt' = h(x,t)\tau$ and $\int_t^{t+\tau} g(x,t')\Gamma(t')dt' = g(x,t)\int_t^{t+\tau}\Gamma(t')dt'$ were used.

For the diffusion coefficient it holds

$$\begin{aligned}
D^{(2)}(x_0,t) &= \frac{1}{2}\lim_{\tau\to 0}\frac{1}{\tau}\langle (X(t+\tau)-X(t))^2 | X(t)=x_0\rangle \\
&= \frac{1}{2}\lim_{\tau\to 0}\frac{1}{\tau}\langle \int_t^{t+\tau}\left[h(X,t')+g(X,t')\Gamma(t')\right]dt' \int_t^{t+\tau}\left[h(X,t'')+g(X,t'')\Gamma(t'')\right]dt'' | X(t)=x_0\rangle \\
&= \frac{1}{2}\lim_{\tau\to 0}\frac{1}{\tau}\langle \int_t^{t+\tau}\int_t^{t+\tau} h(X,t')h(X,t'')dt'dt'' \\
&+ \int_t^{t+\tau}\int_t^{t+\tau} g(X,t')\Gamma(t')g(X,t'')\Gamma(t'')dt'dt'' | X(t)=x_0\rangle \\
&= \frac{1}{2}h(x_0,t)^2\lim_{\tau\to 0}\tau + \frac{1}{2}g(x_0,t)^2\lim_{\tau\to 0}\frac{1}{\tau}\int_t^{t+\tau}\int_t^{t+\tau}\Gamma(t')\Gamma(t'')dt'dt'' | X(t)=x_0\rangle \\
&= \frac{1}{2}g(x_0,t)^2\lim_{\tau\to 0}\frac{1}{\tau}\int_t^{t+\tau}\int_t^{t+\tau} 2\delta(t'-t'')dt'dt'' \\
&= \left(g(x_0,t)\right)^2 . \qquad (2.26)
\end{aligned}$$

Equation (2.24b) is the only result from the theory of stochastic processes summarized in this chapter, where the extension to higher dimensions is not completely in analogue. The arising problematic can already be seen with the transition to two dimensions. In this case, the drift $D^{(1)}(x,t)$ is given by a vector and the diffusion $D^{(2)}(x,t)$ by a matrix. Therefore $\mathbf{g}(x,t)$ is given by a matrix fulfilling $\mathbf{g}(x,t)^T\mathbf{g}(x,t) = D^{(2)}(x,t)$ and thus, the Langevin coefficients are not unique: If $\mathbf{g}(x,t)$ solves equation (2.24b), then all matrices of the form $\tilde{\mathbf{g}} = \mathbf{gO}$ where $\mathbf{O}$ is an orthogonal matrix ($\mathbf{OO}^T = \mathbf{1}$) are also admissible solutions. Usually, the Langevin equation selected is given by the matrix $\mathbf{g}(x,t)$ that is the square root of $D^{(2)}$, computed through an eigenvalue decomposition.

2.7 Stochastic data modeling

The previous sections introduced the two fundamental tools for the description and analysis of stochastic processes, the Langevin and the Fokker-Planck equation. The goal of stochastic data modeling is to extract the coefficients defining these equations directly from synthesized or real time-series data. This is achieved in two steps [Friedrich et al., 2011].

First, to apply the Langevin method it is necessary to ensure that the data is stationary and Markovian. Testing for stationarity is done by measuring the numerical stability of the first two moments of the observed stochastic process. To determine whether the process is Markovian or not it is necessary to estimate the so-called Markovian time scale, which is the minimum time interval over which the data can be considered as a Markov process. For time increments smaller than the Markov scale dynamical noise sources become correlated and also measurement noise can spoil the Markovian properties [Einstein, 1905; Lück et al., 2006].

There are several tests that can be carried out to achieve this, for example the Wilcoxon-test [Wilcoxon, 1945] or checking directly the validity of the Markov property or the Chapman-Kolmogorov equation, for details the reader may refer to Friedrich et al. [2011]. The papers presented in this thesis omit this step and instead create synthetic, Markovian data with the same statistic properties as the original data by the means of a Markov chain model. This is described below and in Chapter 3.

In a second step, the drift and diffusion coefficients are determined. Equations (2.21) and (2.24) show, that any Markov process can be completely described once the drift and diffusion coefficients are known. As already indicated in their definition by Equation (2.15), they are computed as the small τ-limit of the conditional moment [Friedrich and Peinke, 1997]:

$$D^{(i)}(x,t) = \lim_{\tau\to 0}\frac{1}{\tau}M^{(i)}(x,t,\tau), \qquad (2.27)$$

where the conditional moments are given by Equation (2.8).
The computation of the drift and diffusion coefficient as this limit can be difficult for several reasons [Kleinhans, 2008]. First, often there is not enough data to compute reliable results. If there is little data available, the discretization process of the time-series that precedes the computation of the drift and diffusion coefficient leads to little unoccupied bins. Therefore, the averaging in the computation of the conditional moments is done over a small amount of data and therefore the results are not stable.
In this thesis, the problem of having too little data is overcome by analyzing synthetically generated time-series of arbitrary length (see Chapters 3, 4, 5, 6).
Second, each data set has a finite sampling frequency and thus, there is a minimal $\tau = \Delta t$ value for which the fraction in Equation (2.27) can be computed. In many cases however, for the smallest values of τ the conditional moments depend linearly on τ and the drift and diffusion coefficients can therefore be estimated by the quotient between the corresponding conditional moment and τ in this range [Rinn et al., 2016; Siegert et al., 1998].
Once the drift and diffusion coefficients are obtained, they serve to analyze the dynamical behavior of the system under study. The methodology has the advantage, that the coefficients are directly estimated from the data and therefore no *a priori* assumptions about their functional form are necessary. Moreover, it is computationally cheap. It has been applied to various areas, for example in medicine, where Friedrich et al. [2000b] analyzed Parkinson data or Kuusela [2004] revealed pathological cardiac dynamics using a stochastic heart-rate model. In the world of finance, Friedrich et al. [2000a] showed that the U.S. dollar-German mark exchange rates upon different delay times can be analyzed using the Langevin framework. The method has also been applied to traffic flow data [Kriso et al., 2002], road surface profiles [Waechter, M. et al., 2003] and, of course, wind energy. Boettcher et al. [2003] analyzed wind gust measurements near the north sea border of northern Germany and Sura and Gille [2003] interpreted the wind-driven southern ocean variability in a stochastic framework. For a comprehensive review of the applications of the Langevin method, the reader may refer to [Friedrich et al., 2011].
Since all data describing stochastic processes is available as a series of measurements at discrete times, discrete Markov processes are a valuable tool for description and analysis.
To model a dataset using a Markov chain $X(t)$ with a discrete finite state space $\mathscr{S} = \{s_0, s_1, ..., s_n\}$, $n \in \mathbb{N}$ it is necessary to estimate the transition probabilities p_{ij} that describe the transition from state s_i to state s_j for all i, j. This is usually done by maximizing a likelihood function which describes the realization probability of a given dataset [Anderson and Goodman, 1957]. For a sequence of M states $\{X(t_0) = s_{i_0}, ..., X(t_M) = s_{i_M}\}$ with $s_{i_0}, ..., s_{i_M} \in \mathscr{S}$ and $i_0, ..., i_M \in \{0, ..., n\}$, its probability can be computed as $p\{X(t_0) = s_{i_0}\} p_{i_0 i_1} p_{i_1 i_2} \cdot \ldots \cdot p_{i_{M-1} i_M}$. Since the term $Pr\{X(t_0) = s_{i_0}\}$ is constant, given a set of observed state transitions S, it is possible to estimate p by maximizing the likelihood

$$F = \prod_{(ij) \in S} p_{ij}, \tag{2.28}$$

subject to $\sum_j p_{ij} = 1 \forall i$ and $0 \leq p_{ij} \leq 1 \forall i, j$, where a transition is described by an ordered pair (i, j) indicating the origin and the destination of the transition. Introducing the notation n_{ij} for the number of transitions (i, j) in the dataset and yields

$$F = \prod_{i=1}^{n} \prod_{j=1}^{n} p_{ij}^{n_{ij}}, \tag{2.29}$$

which can be simplified to the log-likelihood

$$F_{\log} = \sum_{i=1}^{n} \sum_{j=1}^{n} n_{ij} \log(p_{ij}). \tag{2.30}$$

Using Lagrange multipliers [Ruszczynski, 2006], $\lambda_1, ..., \lambda_n$ to express the row-stochasticity constraint, i.e. the constraint ensuring that all transition probabilities from a state sum up to one, yields the objective function

$$\mathscr{F} = F_{\log} - \sum_{i=1}^{n} \lambda_i \left(\sum_{j=1}^{n} p_{ij} - 1 \right). \tag{2.31}$$

Through setting the derivation with respect to p_{ij} to zero it follows that

$$0 = \frac{n_{ij}}{p_{ij}} - \lambda_i \tag{2.32}$$

and inserting this in the row-stochasticity constraint $\sum_{j=1}^{n} p_{ij} = 1$ yields

$$\sum_{j=1}^{n} \frac{n_{ij}}{\lambda_i} = 1 \Leftrightarrow \lambda_i = \sum_{j=1}^{n} n_{ij}, \tag{2.33}$$

which together with Equation (2.32) provides the maximum likelihood estimator

$$p_{ij} = \frac{n_{ij}}{\sum_{j=1}^{n} n_{ij}} \tag{2.34}$$

for the transition probabilities p_{ij}.
In this thesis, an optimization problem was formulated based on several maximum likelihood estimators in order to include also two and three step transitions in the estimation process. This way, better estimates of the transition probabilities are obtained. Details can be found in Chapter 3.
Once the transition matrix $\mathbf{P}$ is estimated it is possible to generate synthetic data following a Monte Carlo approach. First, the cumulative probability transition matrix $\mathbf{P}_{cum}$ with $P_{cum}(i,j) = \sum_{k=1}^{j} p_{ik}$ is computed and an initial state s_i randomly selected. A random number ε between zero and one is then uniformly selected and a new state s_l is chosen such that $P_{cum}(i,l) \geq \varepsilon$. For details see Sahin and Sen [2001].
For the case of a time-inhomogeneous Markov process the generation process is the same, only for each time t the appropriate transition matrix $\mathbf{P}^{(t)}$ needs to be used.
Discrete, homogeneous Markov chains have widely been used for the generation of synthetic wind data, both wind speed [Shamshad et al., 2005; Sahin and Sen, 2001; Nfaoui et al., 2004], direction [Ettoumi et al., 2003] and power [Papaefthymiou and Klöckl, 2008]. They have also been employed to model turbulent wind speed data [Kantz et al., 2004] and extreme wind speeds [Fawcett and Walshaw, 2006].
Synthetic wind data that includes daily patterns is not as wide-spread and there is no "standard" way to do it. Suomalainen et al. [2013] sampled a probability distribution matrix based on selected daily patterns and the mean wind speed of each day. A different approach was employed by Carapellucci and Giordano [2013] who used a genetic algorithm to estimate the parameters of their model. In this thesis, a new method to generate wind speed, power and direction data, incorporating daily patterns is introduced in the form of a time-inhomogeneous Markov process. The transition probabilities are modeled by Bernstein polynomials to reduce the number of parameters to be estimated from the number of samples per day to the polynomial degree. The details of the model and its implementation are introduced in Chapter 5.

2.8 Empirical data and measurement noise

A problem when dealing with real-world data, i.e. a set of measurements, is that often it is spoiled by measurement noise, which means that an additional noise source, generated by measurement devices for example, is present in the data. This measurement noise is different from the intrinsic dynamical

noise of the stochastic process and does not contain information about the process itself. When analyzing a noisy time-series by the means of stochastic modeling as introduced in Section 2.7 above, instead of the conditional moments only their noisy analogues can be determined and as a result the corresponding drift and diffusion coefficients are distorted.
Several numerical schemes have been proposed to separate measured data into the underlying dynamics and the external measurement noise. A common approach is to optimize a likelihood function that uses Bayesian methods to extract the most probable set of parameters from a given dataset [Meyer and Christensen, 2000, 2001; Fullana and Rossi, 2002; Heald and Stark, 2000]. However, these methods are computationally expensive, since they involve high dimensional integrations over all data points.
In the context of the Langevin analysis Siefert et al. [2003] investigated the case of measurement noise with small amplitudes. They compared the Itô-Taylor expansion with a polynomial that describes the conditional moments for small values of τ and found that the values of drift and diffusion coefficients are the coefficients of the linear term of the expansion. Therefore, they can be used instead of the τ-limit in Equation (2.27).
For the case of strong measurement noise Böttcher et al. [2006] investigated the behavior of the moments for small values of τ. They modeled the measurement noise as Gaussian white noise, assuming it to be memory-less and uncorrelated with the state of the stochastic process. The resulting method is capable of separating the measurement noise from the stochastic process and with a Bayesian *ansatz* they show how to extract the parameters of the stochastic process.
An alternative methodology to tackle data spoiled with Gaussian, delta-correlated measurement noise was proposed by Gottschall and Peinke [2008b], which was then extended by Lind et al. [2010]. Using expansions of the noisy moments they derived a cost function whose minimization yields the amplitude of the measurement noise as well as the parameters of the underlying stochastic process. The method works for measurement noise which has the same order of magnitude as the uncontaminated stochastic signal.
Lehle [2011] introduced an approach that is able to deal with strong exponentially correlated Gaussian measurement noise and subsequently extended it to N dimensions [Lehle, 2013]. Different from previous methods, he started from the joint probability densities of the noisy time-series. Using the assumption that the underlying stochastic process and the measurement noise are independent random variables he expressed the joint probability density of the noisy process as a convolution of the joint probability densities of the uncontaminated stochastic process and the measurement noise, which yield the measurement noise parameters. Using the special properties of the Gauss function with regard to convolution operations, an equation system that relates the uncontaminated moments with their noisy counterparts is derived. Finally, using integral transforms, he extracted polynomial approximations of the drift and diffusion functions. Details of the method can be found in [Lehle, 2013] and the appendix of Chapter 7.
For most of the procedures named above, an *a priori* knowledge of the functional form of the drift and diffusion functions is needed. The method presented in Chapter 7 introduces a parameter-free method that is based on the approach presented by Lehle [2013]. In a more general framework, the algorithm can be considered as a way to separate two stochastic signals.

2.9 Wind energy

Wind energy is a broad and interdisciplinary topic and several books cover the engineering issues related to the production of wind energy [Manwell et al., 2009; Gasch and Twele, 2012; Burton et al., 2001], while others deal with more fundamental questions like turbulence [Davidson, 2004; Hölling et al., 2014].
Assuming a wind turbine with its rotor blades perpendicular to the wind flow, the theoretically available power p from wind speed v can be determined as the kinetic energy $E = \frac{1}{2}mv^2$ of the air

mass m per unit of time t. The mass flow is proportional to the air density ρ and the area of the rotor disc A and given by [Manwell et al., 2009]

$$\frac{dm}{dt} = \rho A v. \tag{2.35}$$

Thus, the power of the wind that flows at velocity v through the area A is given by

$$p = \frac{dE}{dt} = \frac{1}{2}\frac{dm}{dt}v^2 = \frac{1}{2}\rho A v^3. \tag{2.36}$$

This however is not the maximal power a wind energy conversion system (WEC) can extract from the air. The power of the wind is converted into mechanical power of the rotor by slowing down the flowing air mass and thus it cannot be converted completely, since this would stop the moved airmass and thus have the impact of a wall [Peinke et al., 2014; Gasch and Twele, 2012]. Therefore, an optimally working WEC must have a certain wind permeability and wind power output is therefore given as

$$p = \frac{1}{2}\rho A v^3 c_p, \tag{2.37}$$

with $0 < c_p < 1$ being the power coefficient. Betz [1926] proved that the maximum power is extracted if the original upstream wind velocity is reduced to a third of its value downstream the rotor. He showed that for this case of an idealized wind turbine extracting the maximum amount of power from the wind the power coefficient is $c_{Betz} = \frac{16}{27} \sim 0.59$, i.e. only around 60% of the winds power can be converted. Modern wind power plants can reach power coefficients up to 0.5 [Peinke et al., 2014].
The power curve or performance curve is a central concept in wind energy. It shows how the power output of a wind turbine varies with wind speed (at hub height). It is characteristic for every turbine and can normally be obtained by the manufacturer.

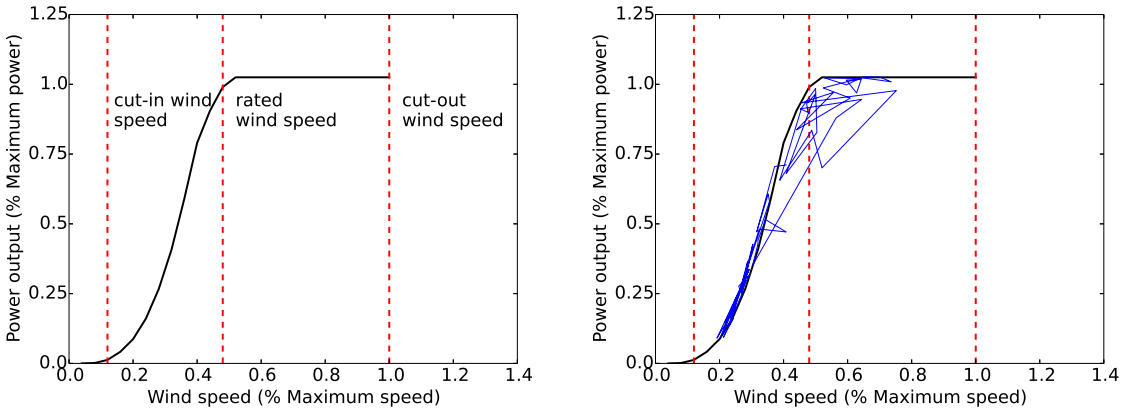


Figure 2.1: Power curve for a wind turbine as specified by the producer and the three key points of wind power productions on the velocity scale (left). Example of a 12-hour trajectory of wind speed vs. power output (right) with a 10-minute sampling rate. For confidentiality reasons, power and speed measurements are given in percentage of their respective maximum value.

Figure 2.1 (left) illustrates the three key points of the power curve on the velocity scale:

- The cut-in wind speed: At this wind speed the turbine begins to rotate and generate power.
- The rated wind speed: At this wind speed the turbine produces the rated power (nameplate power), the maximum level at which it can produce.
- The cut-out wind speed: At this wind speed the turbine stops to deliver power due to safety constraints.

The power curve can only describe the average behavior of the energy conversion process, but not the fluctuations that occur around it, which affect both the power output and the loads on the turbine. On the right-hand side of figure 2.1 an example of a 12-hour trajectory is shown. Even on this 10-minute average data, it can be seen, that the stochastic behavior of the wind source translates to the wind power output. On smaller time-scales, there are stronger fluctuations: Milan et al. [2013] investigated the trajectory of power production on 1 Hz data and observed turbulent fluctuations in both wind speed and power. They argue that this indicates a highly dynamical conversion process and propose to describe it with a dynamical power curve based on the Langevin approach as suggested by Gottschall and Peinke [2008a]. Their proposed method is to compute the power curve as the fixed points of the drift vector field computed by Equation (2.27).
Chapter 4 investigates the application of the Langevin method on data with a sampling rate of 10 minutes and proposes an integration scheme for the determination of the power curve that also takes into account the diffusion coefficients.

2.10 Mean-variance portfolio theory and applications to the energy market

Wind energy is becoming a top contributor due to rather high capacities and generation costs that are becoming competitive with conventional energy sources [Wen et al., 2009]. However, incorporating wind energy into the energy mix poses several challenges, for example concerning the location of the parks, diversification of the energy mix, policy making to encourage investments in wind parks and bidding on the liberalized electricity markets. For all of these topics, the mean-variance portfolio (MVP) or risk-return approach has become a valuable tool to aid decision making. First introduced by Markowitz [1952] in the context of portfolio selection in finance, this approach assesses the ratio between the expected return and the associated risk of return. An investment decision then is a trade-off between risk and expected return, since “the portfolio with maximum expected return is not necessarily the one with minimum variance” [Markowitz, 1952]. This approach is also called mean-variance portfolio since the expected returns can be computed by the expected value or mean and the risk can be expressed by the variance or standard deviation. The higher the variance, the wider the gap between worst and best case scenario of outcome, i.e. the higher the risk of loss.
MVP was first applied to the energy sector by Bar-Lev and Katz [1976] to optimize the fossil fuel mix in the US. Awerbuch et al. [2003] and Awerbuch [2006] were the first to investigate the generation portfolio of the European Union from a risk-return perspective. Similar studies were performed for Switzerland [Madlener and Wenk, 2008; Krey and Zweifel, 2006], the US [Krey and Zweifel, 2006], Italy [Arnesano et al., 2012], Germany [Westner and Madlener, 2011], the Netherlands [Jansen et al., 2006], Portugal [Cunha and Ferreira, 2015], Taiwan [Huang and Wu, 2008], China [Zhu and Fan, 2010] and Brazil [Losekann et al., 2013]. Several of these studies found that the addition of renewable energies [Huang and Wu, 2008; Awerbuch et al., 2003] and specifically wind energy [Awerbuch et al., 2003; Madlener and Wenk, 2008; Doherty et al., 2006] to the existing portfolio can reduce the electricity cost due to fossil fuel price fluctuations.
To reach their targets for electricity production from renewable energy sources, many countries try to increase investment in renewable energy projects [Kitzing, 2014]. To guide investors in the desired direction, it is necessary that appropriate policy schemes are employed that take similar goals into consideration as private investors when preparing investment decisions. Kitzing [2014] analyses two of these support instruments (feed-in tariffs and feed-in premiums) in terms of a mean-variance portfolio analysis.
The major drawback of wind energy systems is the fluctuating nature of their source [Milan et al., 2013]. One way of dealing with this issue is geographic diversification of wind farms to smooth out the fluctuations in wind power generation. Using an MVP approach Roques et al. [2010] analyzed historical wind data from five European countries to define optimal cross-countries wind power

portfolios. First, they optimize the wind power output and, second, maximize the wind power contribution to system reliability. A similar approach was used by Rombauts et al. [2011].
The concept of geographical smoothing of wind power fluctuations is controversial, as there are indications that large power fluctuations can occur on large spatio-temporal scales, on the order of the correlation length of the atmosphere of hundreds of kilometers [Milan et al., 2013]. In any case, the intermittent character of wind energy imposes a challenge to the wind power producer when it comes to reserve commitment and trading on the liberalized electricity market. For that, participants must bid in advance and the uncertainty of wind power production can lead to differences between the committed and actually produced energy [Hosseini-Firouz, 2013]. This imbalance may result in the payment of penalties which decreases the revenue [Pinson et al., 2007]. Therefore, to achieve maximum profit, it is necessary to develop optimal offering strategies. A review on the methods employed for deriving bidding strategies is given by Li et al. [2011].
A MVP-based approach to assess the return of a wind farm in terms of risk that can be used to develop bidding strategies for market participation is introduced in this thesis. Chapter 8 introduces a measure for estimating the best risk-return relation of power production in wind farms within a given time lag conditioned to the velocity field.

Part II

Contributions

Chapter 3

On the use of Markov chain models for the analysis of wind power time-series

The main topic of this thesis is the development of stochastic methods for the analysis and simulation of wind energy production, with a special focus on the application of the Langevin method. This paper presents a Markov Chain model that, on its own, serves to characterize the production behavior of a turbine based on the measurements of wind power, speed and direction. Moreover, a specific application, namely the persistence analysis of power production is suggested and described. In the context of the Lanvegin analysis, the here presented model serves as some sort of a filter for difficulties arising from measured data, such as missing data points or non-stationarity, which would hinder an analysis in the Langevin framework. In contrast to state-of-the-art Markov chain models for the synthesis of wind data, the parameter estimation procedure takes multi-step transitions into account which allows a more efficient use of the usually limited available data, another frequent challenge when dealing with real-world time-series. Moreover, it is based on a joint discretization of wind power, speed and direction data collected from the same turbine and thus it can capture the interdependencies of these variables as well as the dynamic behavior of the conversion process and reproduce them in synthesized data. Moreover, the novel estimation procedure via a nonlinear optimization problem is described in detail and it is shown that the model serves to represent the original data, since all theoretically derived properties are in close agreement with their empirical analogues.

On the use of Markov chain models for the analysis of wind power time-series

Vitor V. Lopes[1], Teresa Scholz[1], Ana Estanqueiro[2], Augusto Q. Novais[1]
[1]Energy Systems Modeling and Optimization Unit
[2]Ocean, Wind and Solar Energy Systems Unit
LNEG, National Laboratory for Energy and Geology
Estrada do Paço do Lumiar, 22, 1649-038, Lisboa, Portugal
vitor.lopes@lneg.pt

Abstract—**Wind energy is becoming a top contributor to the renewable energy mix, which raises potential reliability issues for the grid due to the fluctuating and intermittent nature of its source. This paper explores the use of Markov chain models for the analysis of wind power time-series. The proposed Markov chain model is based on a 2yr dataset collected from a wind turbine located in Portugal. The wind speed, direction and power variables are used to define the states and the transition matrix is determined using a maximum likelihood estimator based on multi-step transition data. The Markov chain model is analyzed by comparing the theoretically derived properties with their empirically determined analogues. Results show that the proposed model is capable of describing the observed statistics, such as wind speed and power probability density as well as the persistence statistics. It is demonstrated how the application of the Markov chain model can be used for the short-term prediction of wind power.**

Keywords-component: Discrete Markov chain models, wind power, variability, persistence

I. Introduction

The EC European Parliament objective to achieve 20% of the consumed energy from the renewable energy sector by 2020, introduced a major challenge to the planning and operation of power systems. Wind energy is becoming a top contributor to the renewable energy mix, which raises potential reliability issues for the grid due to the fluctuating and intermittent nature of its sources. Thus, the understanding of wind speed characteristics and its impact upon the power production is an important task, as shown by the extensive literature addressing wind speed forecast [1] and wind power prediction [2]. Generally speaking, the existent modeling approaches can be classified into: physical, statistical, or a combination of both.

Among the different statistical approaches, Markov chain models are a frequent choice in the published literature and the dominant trend over the years has been the increase on the number of states to improve their performance [3]. The main applications for the Markov models are: a) the short term wind power prediction; and, b) the simulation of wind speed data, however with some limitations in capturing the long term characteristics of the autocorrelation function [4].

This paper addresses the development of a Markov chain model for a wind power turbine using a 2-year historical dataset collected by the device data logger. The main goal is to assess whether Markov chain models can be used to capture statistics commonly used for wind power characterization, such as production level persistence, by comparing the statistics computed from the transition matrix with the ones obtained directly from the data. Moreover, the power prediction capabilities of the model are discussed along with the analysis of the impact of different wind speed uncertainty levels.

The paper first proposes a joint discretization of the wind speed, direction and power variables for the state definition. Then, it presents a multi-step maximum likelihood estimator for the determination of the Markov chain transition matrix. Finally, the results are shown and discussed.

II. Modeling the wind power time-series

A. Wind power data

In this study, the data was obtained from a wind power turbine from a wind park located at the Pinhal Interior region in Portugal. The time-series comprises a two year period (2009/10) of historical data obtained from the turbine data logger, with a 10 minute sampling rate. The wind speed and direction information is collected from the anemometer placed in the wind turbine hub. Due to confidentiality, wind power and speed data values are reported as a fraction of their corresponding maxima.

B. Discrete Markov chain model

A discrete finite Markov process $\{X_k \in S, k \geq 0\}$ is a stochastic process on a discrete finite state space $S = \{s_1, ..., s_n\}$, $n \in \mathbb{N}$ that is characterized by the following property: the probability of a state at any step of the process only depends on the previous state [5]. The Markov property is expressed mathematically by

$$\Pr\{\mathrm{X}_{k+1} = s_j \mid X_k = s_i \wedge X_l \in S \quad \forall l = 0, ..., k-1\} = \Pr\{\mathrm{X}_{k+1} = s_j \mid X_k = s_i\} = p_{ij}(k). \tag{1}$$

978-1-4577-1829-8/12/$26.00 ©2012 IEEE

If the probability of moving to state s_j given that the Markov process is in state s_i, is time-independent, the process is called a Markov chain. Thus, a Markov chain has an associated transition probability matrix $\mathbf{P}$, i.e. a n-by-n matrix with entries $\mathbf{P}(i,j) = p_{ij}$ for all $i, j \in \{1, \ldots, n\}$. $\mathbf{P}$ is a first-order transition matrix and satisfies the following properties: $p_{ij} \geq 0$ and $\sum_j p_{ij} = 1$ for $i, j \in \{1, \ldots, n\}$.

$$\mathbf{p}_{k+1} = \mathbf{p}_k \mathbf{P} \quad \text{with} \quad \mathbf{P} = \begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{pmatrix} \tag{2}$$

The discrete probability vector $\mathbf{p}_{k+1}$, which describes the probability distribution for each state at $k+1$, can be computed from $\mathbf{p}_k$ using (2). Thus, for a given initial state probability vector $\mathbf{p}_0$, it is possible to determine the probability distribution for any given number of time-steps.

C. *Markov chain state definition*

Discrete Markov chain models applied to describe continuous variables require the definition of the states. This work proposes to characterize the wind turbine states using three different variables: wind power, speed and direction. As such, each state will be defined by all the points inside a polyhedron in a three dimensional space.

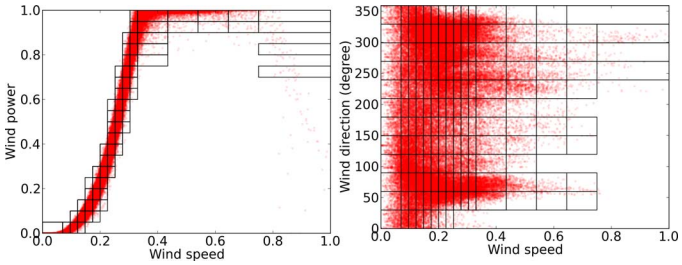


Figure 1. Representation of all data points projected into the: a) wind power and speed plane (left); and, b) wind direction and speed plane (right).

Fig. 1 presents all data observations and the final state definitions projected into: a) the wind direction and speed plane; and, b) the wind power and speed plane, where each rectangle is the projection of a state polyhedron into the two planes. Overall, they define the final state partition for the three-dimensional variable space. As expected, the observations projected into the wind power and speed plane define the characteristic power curve of the wind turbine. In the wind direction and speed plane, data is widely scattered. Three accumulation regions can be identified: one for low wind speeds and two defining the dominant wind directions around 50º and 250-300º.

The data space is discretized unevenly. The wind direction and power are divided by an equally spaced grid leading to 12 and 20 classes, respectively. Due to the different slope characteristics in the power curve, the wind speed is divided as follows. One class for all the values below the cut-in speed and another for all the values above the cut-out speed. In the power curve high slope region (between the cut-in and rated wind speeds), the discretization is narrowed with the selection of 10 classes, whereas for the high power plateau (between the rated and cut-out wind speeds) discretization is widened and 4 classes are defined. The complete state set is constructed through the enumeration of all the possible combinations between the classes of each variable. Due to the relation between the three variables, most of these states are empty. Thus, an initial pruning step is performed to discard those that contain less than 5 observations, which reduces the number of states from 3840 to 573.

Section III.C. describes a preprocessing step required to remove a specific type of state transition, which leads to a further reduction of the number of states. Fig. 1 presents the final result. All the states defined are used in the Markov chain model and analyzed and discussed in section V..

III. Markov chain parameter identification

A. *Maximum likelihood estimator based on multi-step transition data*

A Markov chain process is characterized by its transition matrix $\mathbf{P}$, estimated through the optimization of a constrained maximum likelihood function that describes the probability of the realization present in the dataset [6]. For a sequence of M states $\{X_0 = s_0, ..., X_M = s_M\}$ the probability associated with this path can be computed from equation 2 and is given by $p_{01}p_{12}\ldots p_{M-1M}$. Thus, given a set of observed state transitions, it is possible to estimate $\mathbf{P}$ by maximizing:

$$\text{FO} = \prod_{(i,j)\in\mathcal{S}_1} p_{ij} \tag{3}$$

where $\mathcal{S}_1$ is the set of all the valid 1-step transitions observed in the time-series and a transition is described by an ordered pair (i,j) indicating the origin and the destination of the transition. The maximization is performed with respect to the p_{ij} variables and subject to the following constraints:

$$\sum_j p_{ij} = 1, \quad p_{ij} \geq 0 \tag{4}$$

In practice, instead of maximizing FO it is preferable to maximize the log-likelihood function, i.e. $log(FO)$, since it transforms the original mathematical programming problem into an equivalent one that is convex and, thus, efficiently solvable using state-of-art interior point methods [7].

This standard approach only considers one-step transitions. However, collected data contains information about multiple step transitions, which could be used to improve the transition matrix estimate. In particular, 2-step and 3-step transitions are still tractable algebraically.

To make use of this additional transition information, it is required to modify the actual objective function and add two terms, i.e. the 2-step and 3-step transition likelihood functions.

$$logFO = \sum_{(i,j)\in\mathcal{S}_1} log(p_{ij}) + \sum_{(i,j)\in\mathcal{S}_2} log(p'_{ij}) + \sum_{(i,j)\in\mathcal{S}_3} log(p''_{ij}) \quad (5)$$

where $\mathcal{S}_2$ and $\mathcal{S}_3$ are the sets containing all valid 2-step and 3-step transitions, respectively. The 2-step and 3-step transition probabilities are given by $p'_{ij} = Pr\{X_{k+2} = s_j \mid X_k = s_i\}$ and $p''_{ij} = Pr\{X_{k+3} = s_j \mid X_k = s_i\}$, respectively. The matrix defined by p'_{ij} is the transition matrix for the same Markov chain with a doubled sampling period.

With the Chapman-Kolmogorov equations it is possible to express the 2-step and 3-step transition probabilities in terms of the 1-step probability p_{ij}:

$$p'_{ij} = \sum_k p_{ik}p_{kj}, \quad p''_{ij} = \sum_k \sum_l p_{ik}p_{kl}p_{lj} \quad (6)$$

Note that the final objective function is no longer a convex function with respect to p_{ij} and, thus, the determination of the global optimum can not be guaranteed.

B. *Rigorous optimization: solving the nonconvex optimization problem*

The parameters of the Markov chain model are determined by solving the optimization problem using a rigorous numerical solver. The model was formulated making use of the CASADI computation framework [8] and the optimization was performed by Ipopt, a nonlinear interior-point solver [9], which ensures convergence to the global optimum in the case of convex optimization problems.

The initialization of non-convex optimization problems is crucial for the solver to converge to a high quality optimum (possibly global). Thus, special care was devoted to provide the solver with a good initial guess, which is found by solving the convex optimization problem associated with the 1-step objective function. Due to its convexity, the optimum solution always corresponds to the global optimum. With this initialization, the optimization of the non-convex multi-step objective function was achieved in just 12 iterations.

C. *Transition data preprocessing and validation*

To obtain a Markov chain model, the data samples are converted into states following the state classification guidelines provided earlier. Prior to that, the dataset collected from the wind power data logger must be cleaned to remove any record errors and abnormalities. A state transition is valid and placed in the $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$ sets if and only if the elapsed time between the realization of the two states is correct.

To highlight how the state transition sets $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$ are composed from state data, an example is provided. Consider a state sequence with a missing entry: $\{a, c, b, -, b, a, c\}$. Then, the state transition sets are: $\mathcal{S}_1 = \{(a,c),(c,b),(b,a),(a,c)\}$, $\mathcal{S}_2 = \{(a,b),(b,b),(b,c)\}$ and $\mathcal{S}_3 = \{(c,b),(b,a)\}$. This allows to extract the maximum amount of transition information from the collected data and thus generate better transition matrix estimates.

The collected data is a non-contiguous set in time, so it is necessary to filter the $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$ sets to remove inconsistencies that are not expected to occur. For instance, under normal operating conditions, a wind turbine can not be expected to be producing power permanently, hence there can not be absorbent states in the Markov chain, i.e. states that once reached are impossible to exit. Thus, $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$ are analyzed to identify all the transitions that correspond to one of the following conditions: a) there are no transitions from the destination state; b) there are no transitions to the source state; and, c) the source and destination state are the same and conditions a) and b) are both satisfied. The identified transitions are therefore removed from the sets and the state list simplified by removing empty states.

IV. Markov chain analysis

This section highlights how to extract information from the Markov chain transition matrix for the characterization of wind power variability and persistence.

A. *Stationary probability distribution*

A Markov chain can be classified as irreducible if it is possible to visit any state from any other state and periodic if a given state is returned at regular times. A Markov chain that is both irreducible and aperiodic has a single stationary discrete probability distribution π that satisfies

$$\pi = \pi\mathbf{P}. \quad (7)$$

This distribution is independent of the initial state probability distribution $\mathbf{p}_0$ and provides an indication about the long term behavior of the Markov chain, i.e. it is the limiting distribution $\mathbf{p}_k$ when $k \to \infty$. The stationary distribution is determined by performing an eigenvalue decomposition of the transition matrix. The stationary distribution is the normalized eigenvector that corresponds to the eigenvalue equal to 1. According to the Frobenius-Perron theorem, there is only one single eigenvalue with the largest real value (unity) and its associated eigenvector has only positive entries.

The stationary distribution provides the information to compute the expected value for the first return time $T_{i,i}$, which corresponds to the expected number of steps required to return to a state s_i after it is left. Its value is given by $1/\pi_i$.

B. *Persistence statistics*

The persistence of a given state s_i is related with the number of steps the system consecutively remains at this state. It follows a geometric distribution with expected value $(1 - p_{ii})^{-1}$. The interpretation of the persistence times provides a useful insight on the behavior of the wind power turbine. For instance, in order to identify which states on average provide a more consistent power production.

However, the high number of states present in the Markov chain transition matrix hinders a comprehensive and integrated analysis of the data. To overcome this issue, a more general formula is used to determine the expected time that a Markov chain stays consecutively inside a given subset of states [10]. To compute this estimate, first the states are partitioned into two disjoint subsets: $\mathcal{A}$ containing the states of interest; and $\mathcal{B}$, its complement. Then, the matrix $\mathbf{P}$ is rearranged and divided into the following block structure:

$$\mathrm{P} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}, \tag{8}$$

where the last block of rows and columns corresponds to the states in subset $\mathcal{A}$. The cumulative probability distribution of the persistence (τ) of subset $\mathcal{A}$ can be computed by

$$Pr\{\tau \leq k\} = 1 - \frac{\pi_{\mathcal{A}}(\mathbf{I}-\mathbf{D})\mathbf{D}^{k}\mathbf{1}_{\mathcal{A}}}{\pi_{\mathcal{A}}\mathbf{C}\mathbf{1}_{\mathcal{B}}}, \tag{9}$$

where $\pi_{\mathcal{A}}$ is the stationary probability distribution of the states in subset $\mathcal{A}$ and $\mathbf{1}_{\mathcal{A}}$ is the unit column vector with size equal to $\|\mathcal{A}\|$. The persistence is measured in terms of the number of time samples and the expected value is:

$$E\{\tau\} = \frac{\pi_{\mathcal{A}}\mathbf{1}_{\mathcal{A}}}{\pi_{\mathcal{A}}\mathbf{C}\mathbf{1}_{\mathcal{B}}}. \tag{10}$$

V. Results and Discussion

This section analyzes the main statistical properties derived from the transition matrix of the Markov chain by comparing them with the empirical distributions extracted directly from the data.

A. *Characterization of the wind power based on the stationary probability distribution*

The stationary probability distribution of a regular Markov chain describes the long term properties of the states probability distribution $\mathbf{p}_k$ for $k \to \infty$ and can be directly compared with the empirical data distributions provided that the dataset is sufficiently long.

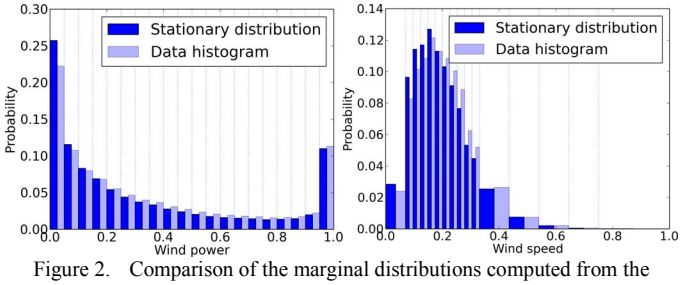


Figure 2. Comparison of the marginal distributions computed from the stationary probability distribution of wind power (left) and wind speed (right) with the respective data histograms.

Figure 2 compares: on the left, the marginal stationary distribution of the power production computed from the Markov chain (dark) with the empirical power distribution obtained from the data (light); and, on the right, the marginal stationary distribution of the wind speed computed from the Markov chain (dark) with the empirical wind speed distribution (light). The respective power and speed states limits are indicated with dotted lines in the background. In general, the two theoretical distributions are in close agreement with the data. The wind power distribution is bimodal, with the two modes located at the minimum and maximum power. The theoretical distribution shows that for this wind turbine the intermediary power levels are relatively rare, for instance, the states corresponding to a power production between 0.4 and 0.9 have a very low probability. The power production distribution is closely linked with the active control strategy defined for the wind turbine and the wind site characteristics. The wind speed distribution follows the expected behavior, a single modal distribution with a long tail for the high wind speeds.

From the stationary distribution of the proposed Markov chain model it is also possible to compute the 2-dimensional marginal probability distribution for the wind power and direction. This provides useful information regarding the dominant wind directions for wind production in this site.

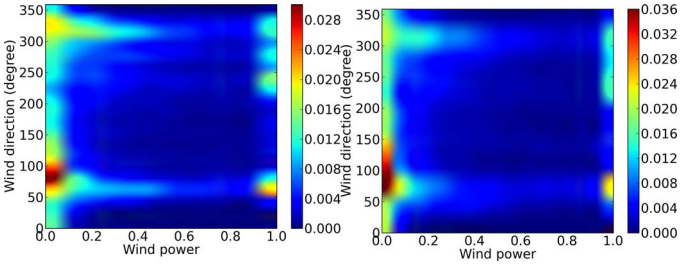


Figure 3. Contour plots: (left) 2D histogram of the wind power and direction data; (right) 2D marginal probability distribution of the wind power and direction computed from the stationary probability distribution.

Figure 3 shows two contour plots: on the left, the empirical 2D distribution of the wind power and direction computed from the data and, on the right, the 2D marginal theoretical distribution computed from the Markov chain's stationary distribution of the states. The comparison of these two distributions shows that the model is capable of capturing the wind power and directions existent in the data. It is possible to see the two dominant directions associated with high wind power production already mentioned in section II.C.. Fig. 3 also provides useful information regarding the direction spread for the high production states, the spread being greater for the 230-330º directions and less for 50-100º. There a clear change from high production states to no production when the wind direction changes from this sector to the 100º-150º sector. The probability for wind production when the wind direction is significantly lower inside the 100º-230º sector.

Figures 2 and 3 clearly demonstrate the capability of this Markov model to capture the combined characteristics of the wind power, speed and direction. The long term behavior of the model is close to what is observed in the dataset.

B. *Characterization of the wind power production persistence*

Persistence statistics allow to characterize the duration distribution of a given state or set of states. In the context of wind power production, persistence allows to characterize the duration of the power production at a certain level. It is used for the analysis of the dynamic characteristics of the Markov chain by comparing the theoretical expected values with the ones computed directly from the data.

Figure 4a, compares the expected values of the persistence for each power state computed from the Markov chain transition matrix (dark) using equation (10) with the values computed directly from the data (light). A close agreement is obtained, indicating that the Markov chain also is capable to capture the information about the dominant system dynamics. The two most persistent states are the no production state and the high power production state, with an expected value of 120 minutes and 90 minutes respectively. This means that, for instance, when the wind turbine enters the high production state it is expected to remain in that state for 90 minutes. Another feature of this wind turbine, is the low persistence observed in the intermediate power states, revealing the action of the wind turbine power controller.

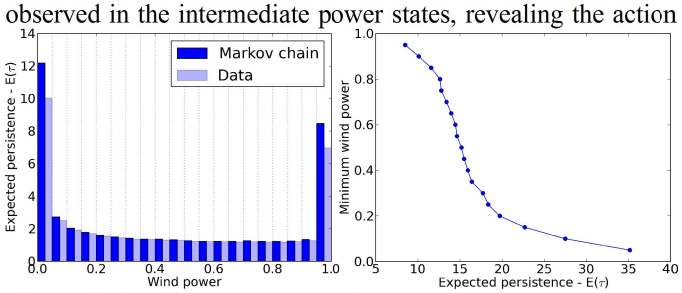


Figure 4. Expected value of the persistence of: a) each power class computed from the data and from the Markov chain transition matrix (left); and b) of the power production to be above a defined threshold (right).

The right side of Fig. 4 shows the expected value of the persistence of the power production to be above a certain power level. This value is computed using equation 10 and defining the set $\mathcal{A}$ as all the states corresponding to a power production not less than a predefined level, i.e. the lower limit of a given power class. The procedure is repeated and the computed values plotted. The upper part of curve (power production greater than 0.3) can be divided into two different sections, with a clear breakpoint at around 80% of the maximum power production. The average slopes in these two regions are quite different and are attributed to the action of the wind power controller. In the first region (down to 80%) the expected persistence changes from around 1.5 hr to 2 hr, whereas for the next region there is an abrupt change in the minimum power production level for the same increase in persistence (0.5 hr). The analysis allows to characterize the capability of the wind turbine to provide a sustainable minimum power level during a given amount of time.

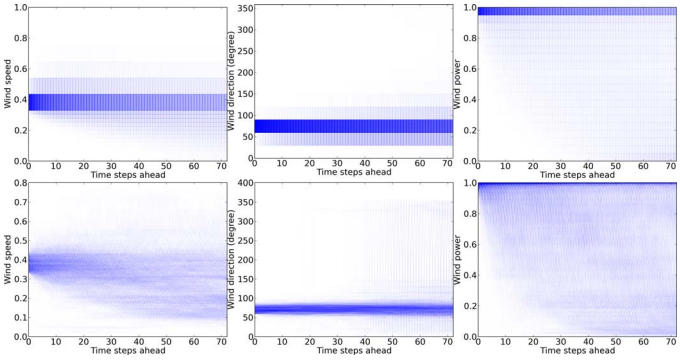


Figure 5. Comparison of the temporal evolution of: (top) the Markov chain wind speed (left), direction (middle) and power (right) probability distribution; (bottom) the wind speed (left), direction (middle) and power (right) for all time-series data segments with the same initial state.

C. *Wind power short-term prediction: value of information*

The Markov chain model is of stochastic nature, meaning that its short-term prediction is expressed in terms of the temporal evolution of the states probability distribution.

The upper row of Fig. 5 shows the temporal evolution of the wind speed, direction and power probability distribution propagated 72 time steps ahead (12 hrs) from a given initial known state. The lower row presents all the 12 hrs sequences present in the dataset that start at the same state. In both row sequences, the intensity level is associated to the probability value, with darker areas reflecting higher probability. The analysis of the two rows shows a good pattern agreement, thus providing additional confirmation that the proposed Markov chain is capable of describing the observed future probability distributions. However, it also highlights the inherent difficulty in providing more informative probability distributions for the future wind power state, since the probability distribution is not confined to a small number of possible states. Note the very different nature of the three variables, with wind speed and direction providing more information regarding its location in the respective state space.

Given that Markov chains provide the probability distribution of future states, it is possible to assess the impact of providing *a priori* knowledge about the future distribution of wind speed states. Broadly speaking, the procedure can be outlined as follows: a) given $\mathbf{p}_k$ compute the future probability distribution $\mathbf{p}_{k+1}$ using equation 2; b) use Bayes' theorem to correct $\mathbf{p}_{k+1}$ given a probability distribution for the wind speed at step $k+1$; and, c) use this corrected probability distribution for the next iteration, i.e. step a. For each step, the wind power distribution is computed from the corrected $\mathbf{p}_k$ marginalizing over the wind speed and direction.

Figure 6. Impact of wind speed prediction at different levels and different accuracies, upon the prediction of the wind power probability distribution. On the left, three different wind speed estimation scenarios. On the right, the correspondent results for the probability distribution for the wind power.

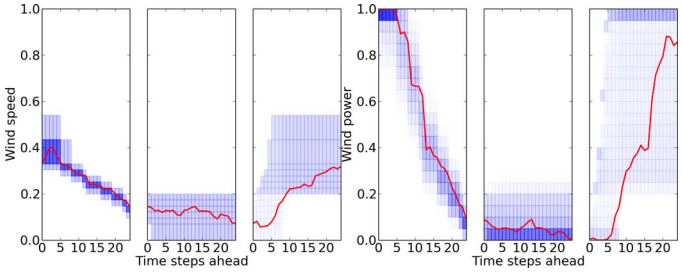


Figure 6 shows the results for three different simulation scenarios. On the left three different cases are presented for the wind speed distribution: 1) based on knowledge provided by an unbiased wind speed estimator; 2) given by a uniform distribution located for low wind speeds; and, 3) a transition between a low wind speed and mid-ranged wind speeds, modeled by a two stage uniform distribution for the first 8 time-steps and a uniform distribution afterwards. On the right, are presented results for the wind power distributions. In Fig. 6, the solid line represents the measured wind speed (left) and the wind power (right). The shaded area represents the probability distribution for each variable, with a darker area attributed to the higher probability values.

As expected, the *a priori* knowledge of the wind speed distribution allows for a better prediction of the distribution of future power states. The analysis of these three cases shows the inherent limitation associated with the wind power prediction. For useful information about the wind power production to be obtained when the wind speed is located around 20 to 40 percent of the maximum wind speed, its *a priori* probability distribution must be sufficiently narrow, spanning over 3 to 4 states. Otherwise, the wind power probability distribution predicted from the Markov chain model will be uninformative, with most of the power states exhibiting a relevant probability value.

VI. Conclusions

This paper explores the use of Markov chain models for the analysis of wind power time-series. Models are developed using the states that combine information about the wind speed, direction and power variables, which are commonly recorded by the data logger from the wind turbine. The joint segmentation of the 3-dimensional variable space allows to decrease the number of the model states, while, simultaneously, encoding the wind power curve into the Markov chain model.

The estimation of the transition matrix is performed by maximizing the maximum likelihood defined using multi-step transition data which allows to extract more information of the transitions in the dataset. Results demonstrate that the proposed Markov chain model can describe satisfactorily the statistical properties of the wind power time-series. It is thus found that the theoretically derived properties from the Markov chain are in a close agreement with the empirically computed analogues.

The use of this Markov chain model for short-term wind power prediction has a performance similar to the standard persistence model, when used in isolation. Markov chain models can be used as a framework to combine information from additional sources. Using Bayes's theorem, it was demonstrated how the wind power probability prediction can be improved by incorporating information about the wind speed distribution. Results show that for an informative prediction of the probability distribution for future states a good prediction for the wind speed is required, in particular if its value is in the high slope area of the wind power curve.

Acknowledgment

The authors gratefully acknowledge the FCT (Fundação para a Ciência e Tecnologia) for the partial financial support for this study (project PTDC/SEN-ENR/114178/2009).

References

[1] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, and Z. Yan, "A review on the forecasting of wind speed and generated power," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 4, pp. 915-920, May 2009.

[2] a Costa, a Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa, "A review on the young history of the wind power short-term prediction," *Renewable and Sustainable Energy Reviews*, vol. 12, no. 6, pp. 1725-1744, Aug. 2008.

[3] A. Carpinone, R. Langella, A. Testa, and M. Giorgio, "Very short-term probabilistic wind power forecasting based on Markov chain models," in *Probabilistic Methods Applied to Power Systems (PMAPS), 2010 IEEE 11th International Conference on*, 2010, pp. 107–112.

[4] K. Brokish and J. Kirtley, "Pitfalls of modeling wind power using Markov chains," in *Power Systems Conference and Exposition, 2009. PSCE'09. IEEE/PES*, 2009, pp. 1–6.

[5] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*, 2nd ed. New York: Springer, 1976, p. 244.

[6] T. W. Anderson and L. A. Goodman, "Statistical Inference about Markov Chains," *The Annals of Mathematical Statistics*, vol. 28, no. 1, pp. 89-110, Mar. 1957.

[7] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge Univ Pr, 2004, p. 730.

[8] J. Andersson and B. Houska, "Towards a Computer Algebra System with Automatic Differentiation for use with Object-Oriented modelling languages," *Object-Oriented Modeling Languages*, pp. 99-105, 2010.

[9] A. Wächter and L. T. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Mathematical Programming*, vol. 106, no. 1, pp. 25-57, Apr. 2005.

[10] K. Anastasiou and C. Tsekos, "Persistence statistics of marine environmental parameters from Markov theory, Part 1: analysis in discrete time," *Applied Ocean Research*, vol. 18, no. 4, pp. 187-199, Aug. 1996.

Chapter 4

Uncovering wind turbine properties through two dimensional stochastic modeling of wind dynamics

The previous chapter introduced a Markov Chain model to generate wind power, speed and direction data. Analyzing these synthetic time-series in the Langevin framework instead of the original measurements overcomes some of the well-known challenges imposed by real-world data, as they are by construction Markovian, stationary, complete and of arbitrary length.
However, an issue that remains is the computation of the τ-limit in the estimation procedure of the drift and diffusion coefficients (see Chapter 2). Often it is argued, that the limit of τ tending to zero in Equation 2.27 can only be estimated when the underlying stochastic process is sampled with a very high frequency. This paper demonstrates, that even if 10-minute average data is used, a common sample frequency available in wind datasets, it is possible to analyze the turbines dynamic behavior in terms of the Langevin approach.
After describing the estimation procedure for the drift and diffusion coefficients, separating the turbines behavior into a deterministic and a stochastic part, the resulting data is analyzed in the wind-speed plane. This allows to uncover specific features of the wind turbine and to characterize its different working regions. Taking into account both drift and diffusion coefficients it is also shown how the characteristic power curve of the wind turbine can be derived.

PHYSICAL REVIEW E **88**, 042146 (2013)

Uncovering wind turbine properties through two-dimensional stochastic modeling of wind dynamics

Frank Raischel,[1,2] Teresa Scholz,[3,4] Vitor V. Lopes,[3] and Pedro G. Lind[1,5]

[1]*Center for Theoretical and Computational Physics, University of Lisbon, Avenida Professor Gama Pinto 2, 1649-003 Lisbon, Portugal*

[2]*Center for Geophysics, IDL, University of Lisbon, 1749-016 Lisbon, Portugal*

[3]*Energy Systems Modeling and Optimization Unit, National Laboratory for Energy and Geology, Estrada do Paço do Lumiar 22, 1649-038 Lisbon, Portugal*

[4]*Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, 1649-003 Lisboa, Portugal*

[5]*ForWind—Center for Wind Energy Research, Institute of Physics, Carl-von-Ossietzky University of Oldenburg, DE-26111 Oldenburg, Germany*

(Received 26 October 2012; revised manuscript received 31 May 2013; published 29 October 2013)

Using a method for stochastic data analysis borrowed from statistical physics, we analyze synthetic data from a Markov chain model that reproduces measurements of wind speed and power production in a wind park in Portugal. We show that our analysis retrieves indeed the power performance curve, which yields the relationship between wind speed and power production, and we discuss how this procedure can be extended for extracting unknown functional relationships between pairs of physical variables in general. We also show how specific features, such as the rated speed of the wind turbine or the descriptive wind speed statistics, can be related to the equations describing the evolution of power production and wind speed at single wind turbines.

DOI: 10.1103/PhysRevE.88.042146 PACS number(s): 02.50.Ga, 02.50.Ey, 92.70.Gt

I. INTRODUCTION

The use of efficient and clean renewable energy sources is one of the major conditions required to achieve the important aim of sustainable development in modern societies [1]. Wind energy is one such source and wind turbines are being subject to intensive studies for improving their efficiency [2,3]. Although the basic laws of atmospheric wind motion have been known for a long time, important problems such as turbulence, layering, and the statistics of extreme events remain poorly understood. A better understanding of these phenomena can help in the construction of energy conversion schemes that are both more efficient and robust. Here two aspects of robustness must be considered. First, the occurrence of sudden changes in wind speed and direction can interrupt the process of energy conversion, meaning unreliability and a sudden slump in the electrical energy generated, which is seen as one of the major obstacles for the replacement of fossil and nuclear plants by wind energy sources. Second, these sudden changes introduce massive mechanical stresses that can lead to excessive wear or, ultimately, to the destruction of wind generators.

Wind flow is in general turbulent [4] and nonhomogeneous [5] with a non-negligible stochastic contribution. Therefore, in order to be able to construct more accurate models for its physical properties, one needs either accurate measurements of the wind speed on the length scales of wind turbines or suitable models that can statistically reproduce these measured data. Since the wind turbines are driven by turbulent wind fields, the stochasticity of the wind fields transfers to the stochastic dynamics of the wind turbine as a whole, of the loads on its structures, and of the power output. Recently, a Markov chain model was used to reproduce wind measured data [6], based on the transition matrix and time propagators for the wind speed and direction together with the power production. Differently from previous first-order approaches [7], information from two- and three-step transition probabilities are considered.

In this paper we aim at understanding the stochastic aspects of power production coupled to the wind velocity field. To that end, we use a methodology introduced in Ref. [8], for uncovering optimal stochastic variables weakly [9] and strongly coupled [8], and adapt it with two purposes: first, to properly derive the functional relation of pairs of variables whose values are extracted from the Markov chain model for wind turbines and second, to uncover specific features of the wind turbine and characterize the different working regions observed in the power-speed plane. After this, we test our approach to uncover the functional dependence of the well-known performance curve, which describes the functional dependence of the power production and the wind speed. Whereas previous reports have pointed out the benefits of deriving the power curve from the drift field [10], we additionally take the diffusion field into account and find that this procedure creates additional insight.

We use the data sets generated by the Markov chain model described by Lopes *et al.* [6]. Using such synthetic data sets that properly reproduce the statistical features of empirical data sets allows us to use data sets as large as needed for our analysis. Moreover, the Markov chain model serves as a filter to remove periodicities present in the data. This is an essential step for our modeling, as it would not be possible to correctly estimate the Kramers-Moyal coefficients from the raw data.

We start in Sec. II by describing the empirical data used to define the Markov chain model as well as the data generated with it. In Sec. III we describe our stochastic method for analyzing the data and in Sec. IV we apply it to analyze the performance curve of a wind turbine. In Sec. V we show that the same method, when applied separately to both wind speed and power production, allows us to derive the performance curve. Further, the same analysis also provides insight concerning specific features of the turbine system studied. Section VI presents a discussion and conclusions.

II. PROPERTIES AND GENERATION OF THE DATA SETS

The data analyzed in this paper were simulated from a set of measurements from a wind turbine in the region of

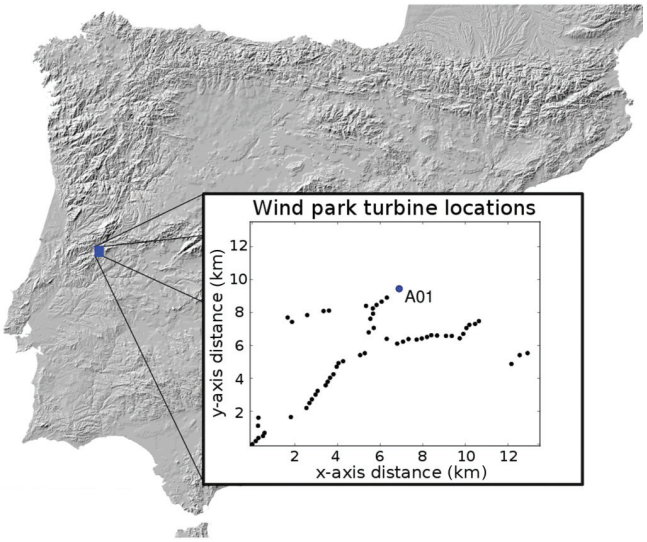


FIG. 1. (Color online) Illustration of the Iberia Peninsula indicating the position of the Portuguese wind park. The inset shows the geographic location of each of the 57 wind turbines (bullets). The blue marker (AO1) indicates the wind turbine analyzed there.

Pinhal Interior, Portugal. The measured properties are the power production P of the wind turbine, the wind speed v, and the wind direction θ ($\theta = 0$ corresponds to north). The wind turbine was selected out of a total of 57 wind turbines in an eolic park. Figure 1 shows an overview of the eolic park. The time increment between two successive measures is $\Delta t = 10$ minutes and the time period covered starts 1 January 2009 and ends 31 December 2010, yielding approximately 10^5 data points. It has to be remarked that these measurements are acquired directly from the top of the wind turbine (nacelle) and might not be optimal for the reconstruction of the underlying physical processes for two reasons. First, the wind speed measurement is acquired at a point located downstream of the turbine blades and cannot account for either the spatial extension and inhomogeneity of the wind field or its complex aerodynamical interaction with the turbine blades [11]. Second, the 10-min sampling period of the historical data set does not allow us to resolve the time scales of either the turbulent interaction between wind and turbine or the quick action of the controller system response. Finally, missing data records, a low number of data points [12] or large sampling intervals [13], and periodicities due to the daily and seasonal variations in wind flow often hinder a direct stochastic analysis of these data sets. Specifically, it is well known that the estimation of the drift and diffusion coefficients that we use cannot be applied to periodic time series. If applicable, a filtering or detrending procedure has to be applied to the data set [9,14,15]. However, as the quasidaily variations in the wind speed do not occur every day, the usefulness of a filtering or detrending procedure has to be doubted. Thus the challenge is to devise alternative methods that can make the most use of the information present in these data sets—given that most of the data acquisition systems on existent wind farms are limited—with the aim of understanding the dynamic processes of the wind power generation, which hopefully can lead to economic benefits from scheduling and maintaining a level of constant production.

To overcome some of these problems, we employ a reconstruction of the original wind data set through a Markov Chain Model (MCM), which has recently been established [6] using a joint discretization of the wind speed, power, and direction variables for the state definition. In our specific case within the range of each variable P, v, and θ we select 80, 60, and 12 states. States without a realization in the time series are deleted and all states that contain at least one observation are kept. Details about the estimation of the transition probabilities can be found in Refs. [6,16] and Appendix A. The use of the MCM has the additional benefit of removing periodicities from the data.

Based on the Markov chain transition matrix $\mathbf{P}$ with $\mathbf{P}(i,j) = p_{i,j}$ being the probability of transition from state s_i to state s_j, the synthetic data sets were generated using the following Monte Carlo approach. We find the cumulative probability transition matrix $\mathbf{P}_{\text{cum}}$ with $\mathbf{P}_{\text{cum}}(i,j) = \sum_{k=1}^{j} p_{i,k}$ and select randomly an initial state s_i. A random number ϵ between zero and one is then uniformly selected and a new state $s_{\mathcal{N}}$ is chosen such that $P_{\text{cum}}(i,\mathcal{N}) \geqslant \epsilon$. (For details see Ref. [7].) Figure 2 shows the generated time series. As shown in Ref. [6], the proposed Markov chain model reproduces the dominant statistical features of all three properties, namely, power production, wind speed, and wind direction, although no periodicities are present in the reproduced time series.

The resulting synthetic data series for power production and wind speed also retain the persistence statistics, namely, the average duration of power production and wind speed on a certain level, respectively. Power production and wind speed are presented as fractions of the maximum observed power P_{max} and wind speed v_{max}, respectively, assuming, therefore, values between zero and one.

The wind direction in the original time series follows a bimodal pattern, which is, to some extent, also periodic: During the day there are weaker breezes in a particular direction than during the night when wind streams in a different direction; therefore, the prevailing wind speed values occurring during the "day" (2 p.m. to 2 a.m.) have approximately $\theta_1 \sim 80^\circ$, a value different from the one found for wind speeds measured during the "night" (2 a.m. to 2 p.m.), $\theta_2 \sim 320^\circ$. The Markov chain model is capable of preserving this bimodality. Although we will not consider the wind direction in our study, one should notice that the wind direction bimodality is reflected in the distribution of the wind speed (see the inset of Fig. 5). Another bimodality preserved from the original data set is the one in the joint probability density function (PDF) of velocity and power, which is visible in Fig. 4(f).

Figure 3 shows the discretization of the data set where the combined states for power and speed are indicated with boxes and circles represent the historical data set points projected into the speed-power plane. The synthetic data sets were generated with 2×10^6 data points and show stationary behavior, i.e., have constant moving averages (not shown).

This approach has several advantages over the direct analysis of historical data sets. First, high-quality data series of arbitrary length can be generated, which increases the accuracy of the Markov analysis. Second, the generated data are by construction Markovian, with the reconstruction through the

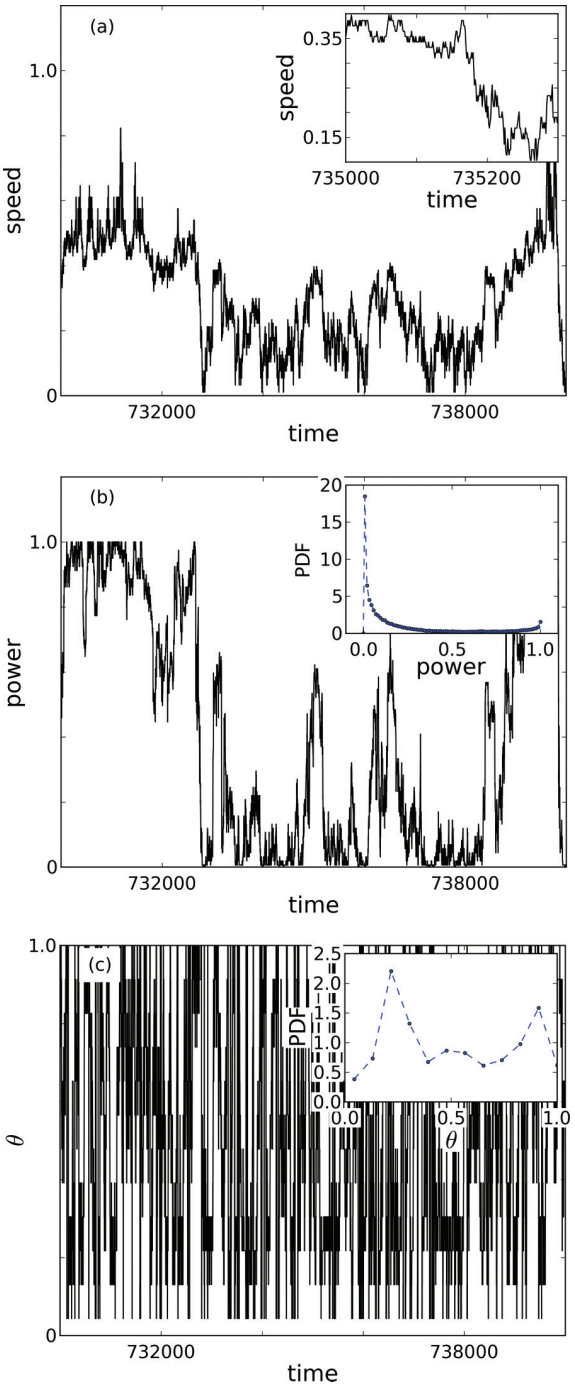


FIG. 2. (Color online) Time series for (a) the magnitude of the wind speed v. The inset shows the time series for a shorter time period. The PDF of the wind speed is shown in the inset of Fig. 5. (b) Power production P of the wind turbine. The inset shows the PDF of the time series. (c) Corresponding wind direction θ. The inset shows the PDF of the time series. All data series were generated with the Markov chain model [6] described in Sec. II. All properties are normalized to the observed intervals $[0, v_{\max}]$, $[0, P_{\max}]$, and $[0, \theta_{\max}]$, respectively. In this and all following figures, multiples of 10 min, whereas velocity v, power P, and direction θ are normalized to unity.

Markov chain acting as a filter that removes both noise correlations and periodicities. Finally, non-Gaussian transition

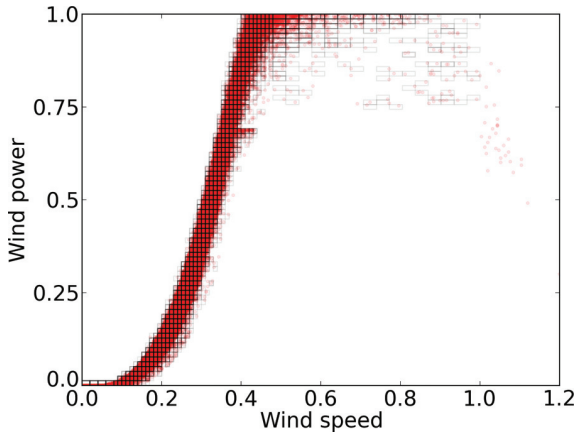


FIG. 3. (Color online) Performance curve for one wind turbine in Pinhal Interior, Portugal. Circles show all the historical data points used in the Markov chain modeling [6] and the boxes display the state discretization.

probabilities between the states are preserved, which enables us to study them through higher Kramers-Moyal coefficients. Error analysis for the derivation of the transition matrix is described in Appendix B.

III. STOCHASTIC ANALYSIS OF WIND TURBINES

The coevolution of two or more stochastic variables, such as wind speed and power production, can be described through a system of coupled stochastic equations, each one defined by a deterministic contribution (drift) and stochastic fluctuations from possible stochastic sources. In this section we present the general framework to analyze our data and in the next section we apply it to the power production and wind speed variables.

For the general case of K stochastic variables $X_1, \ldots, X_K$, the vector $\mathbf{X}(t) = [X_1(t), \ldots, X_K(t)]$ defines the state of the system under study at each time instant t. The evolution of the state vector yields a stochastic trajectory in phase space and is given by the so-called Itô-Langevin equation [8,14,17]

$$\frac{d\mathbf{X}}{dt} = \mathbf{h}(\mathbf{X}) + \mathbf{g}(\mathbf{X})\mathbf{\Gamma}(t), \tag{1}$$

where $\mathbf{\Gamma(t)} = [\Gamma_1(t), \ldots, \Gamma_K(t)]$ is a set of K independent stochastic forces with Gaussian distribution fulfilling the following conditions: $\langle \Gamma_i(t) \rangle = 0$ and $\langle \Gamma_i(t)\Gamma_j(t') \rangle = 2\delta_{ij}\delta(t - t')$. The function $\mathbf{h} = \{h_i\}$ in Eq. (1) is the deterministic contribution, describing the physical forces that drive the system, while $\mathbf{g} = \{g_{ij}\}$ describes the amplitude of the stochastic sources of fluctuations $\mathbf{\Gamma}$ [18].

The evolution of the stochastic variables in time yields a joint PDF $f(\mathbf{X})$ that evolves according to the so-called Fokker-Planck equation

$$\frac{\partial f(\mathbf{X},t)}{\partial t} = -\sum_{i=1}^{N} \frac{\partial}{\partial x_i}\left[D_i^{(1)}(\mathbf{X}) f(\mathbf{X},t)\right] + \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\partial^2}{\partial x_i \partial x_j}\left[D_{ij}^{(2)}(\mathbf{X}) f(\mathbf{X},t)\right], \tag{2}$$

where the functions $D_i^{(1)}$ and $D_{ij}^{(2)}$ are related to the functions h_i and g_{ij} above, namely,

$$D_i^{(1)}(\mathbf{X}) = h_i(\mathbf{X}), \tag{3a}$$

$$D_{ij}^{(2)}(\mathbf{X}) = \sum_{k=1}^{N} g_{ik}(\mathbf{X}) g_{jk}(\mathbf{X}), \tag{3b}$$

and are usually called drift and diffusion functions, respectively. Drift and diffusion functions can be directly derived from observed or generated data [18,19] and this fact is the basis of our framework. Indeed, the drift and diffusion functions of the underlying process are defined through conditional moments, namely [14],

$$\mathbf{D}^{(k)}(\mathbf{X}) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \frac{\mathbf{M}^{(k)}(\mathbf{X}, \Delta t)}{k!}, \tag{4}$$

where $\mathbf{M}^{(k)}$ are the first and second conditional moments ($k = 1,2$). These conditional moments can be directly derived from the measured data as [18,19]

$$M_i^{(1)}(\mathbf{X}, \Delta t) = \langle Y_i(t + \Delta t) - Y_i(t) | \mathbf{Y}(t) = \mathbf{X} \rangle$$

and

$$M_{ij}^{(2)}(\mathbf{X}, \Delta t) = \langle [Y_i(t + \Delta t) - Y_i(t)][Y_j(t + \Delta t) - Y_j(t)] | \mathbf{Y}(t) = \mathbf{X} \rangle,$$

where $\mathbf{Y}(t) = [Y_1(t), \ldots, Y_N(t)]$ is the N-dimensional vector of measured variables and $\langle \cdot | \mathbf{Y}(t) = \mathbf{X} \rangle$ symbolizes a conditional averaging over the entire measurement period, where only measurements with $\mathbf{Y}(t) = \mathbf{X}$ are taken into account. Important conditions to hold are that (i) the underlying process is stationary and (ii) the Markovian property is fulfilled.

Numerically, $\mathbf{h}$ and $\mathbf{g}$ are determined on an $n_1 \times \cdots \times n_N$ mesh of points in phase space, as a function of the variables X_i, using the drift and diffusion functions. Locally, at each mesh point, one can always diagonalize the matrix $\mathbf{g}(\mathbf{X})$ and compute their K eigenvalues and K eigenvectors. As shown previously [8,15,20], this analysis provides information about the stochastic forces acting on the system. Namely, the eigenvalues indicate the amplitude of the stochastic force and the corresponding eigenvector indicates the direction toward which such force acts. In a previous work [8] we argued that to each eigenvector of the diffusion matrix one can associate one independent source of stochastic forcing Γ_i and thus the eigenvectors can be regarded as defining principal axes for stochastic dynamics. In particular, if one eigenvalue is very small compared to all the others, the corresponding stochastic force can be neglected. In the following sections we present a different implication of this principal stochastic component analysis, which emphasizes that the vanishing of one stochastic direction is in fact an indication of a strong functional dependence between the pair of variables being analyzed.

IV. WIND TURBINE DRIFT AND DIFFUSION MAP ANALYSIS

In this section we focus solely on two variables: power production P and wind speed v. Since both series are stationary and Markovian, we assume them to evolve according to the following equations:

$$\frac{dv}{dt} = h_v(v,P) + g_{vv}(v,P)\Gamma_1 + g_{vP}(v,P)\Gamma_2, \tag{5a}$$

$$\frac{dP}{dt} = h_P(v,P) + g_{Pv}(v,P)\Gamma_1 + g_{PP}(v,P)\Gamma_2. \tag{5b}$$

In general, the six functions defining the vector $\mathbf{h}$ and matrix $\mathbf{g}$ depend on both variables and describe the coupling between each other. Based on Eqs. (3), we can derive both $\mathbf{h}$ and $\mathbf{g}$ from the functions $\mathbf{D}^{(1)}$ and $\mathbf{D}^{(2)}$, which in turn are extracted directly from the synthetic data set by computing the corresponding conditional moments using Eq. (4). Note that solving Eq. (3b) for computing the matrix $\mathbf{g}$ yields multiple solutions. If $\mathbf{g}$ is a solution then all matrices of the form $\tilde{\mathbf{g}} = \mathbf{gO}$, where $\mathbf{O}$ is an orthogonal matrix ($\mathbf{OO}^T = \mathbf{1}$), are also admissible solutions. The matrix $\mathbf{g}$ can therefore be computed as the square root of matrix $\mathbf{D}^{(2)}$, i.e., by diagonalizing $\mathbf{D}^{(2)}$ through a proper permutation matrix and, since all eigenvalues are positive ($\mathbf{D}^{(2)}$ is positive definite), taking the square root of each eigenvalue and transforming the matrix back.

Figures 4(a)–4(e) show the five components of $\mathbf{D}^{(1)}$ and $\mathbf{D}^{(2)}$, i.e., the numeric results for both the drift and the diffusion coefficients computed directly from the generated P and v time series. The large fluctuations in the region near maximum power production and wind speed are due to a lack of observations. Indeed, the joint PDF for P and v [Fig. 4(f)] shows that this region is poorly sampled.

To extract valuable information, next we treat these functions separately. Namely, we consider the drift vector field (h_v, h_P) and the eigenvectors of the diffusion matrix associated with its eigenvalues $\lambda_{\max}$ and $\lambda_{\min}$. Figure 5 shows the drift vector field in the power production and wind speed state space, restricted to the sampled region defined by the power production curve in Fig. 3. The solid black line is the performance curve computed from the (P,v) joint probability density function, shown in Fig. 4(f), and defines the most likely power production for a given wind speed. Three different regions can be identified.

Region I in Fig. 5 is characterized by a high wind speed, i.e., above a threshold velocity v_{th} that exceeds the rated wind speed of the turbine. Postponing a more detailed description to Sec. V, we define $v_{\text{th}} = 0.55 v_{\max}$, which corresponds to the 97th percentile of the wind speed distribution. For these rare events of high wind speeds, the expected behavior of the wind turbine is to maintain the power production since there is a surplus of energy in the airflow. In this region, the performance curve is roughly constant at $\sim 0.95 P_{\max}$. Still, positive power drifts are observed whenever the power production is below the performance curve. The wind speed drift is large in magnitude and always negative, i.e., the drift points towards lower wind velocities.

Region II is characterized by production levels above a transition region of $0.6 P_{\max} \lesssim P \lesssim 0.8 P_{\max}$, indicated by gray shading in Figs. 5 and 6, which is defined in more detail in Sec. V and Fig. 6. A closer look at region II enables one to identify a fixed point region ($\mathbf{D}^{(1)} \sim 0$) at high power production levels and wind speed $v \sim 0.5 v_{\max}$, shown as an encircled area in Fig. 5 near the v_{th}. This speed value coincides approximately with the rated wind speed, i.e., the speed for

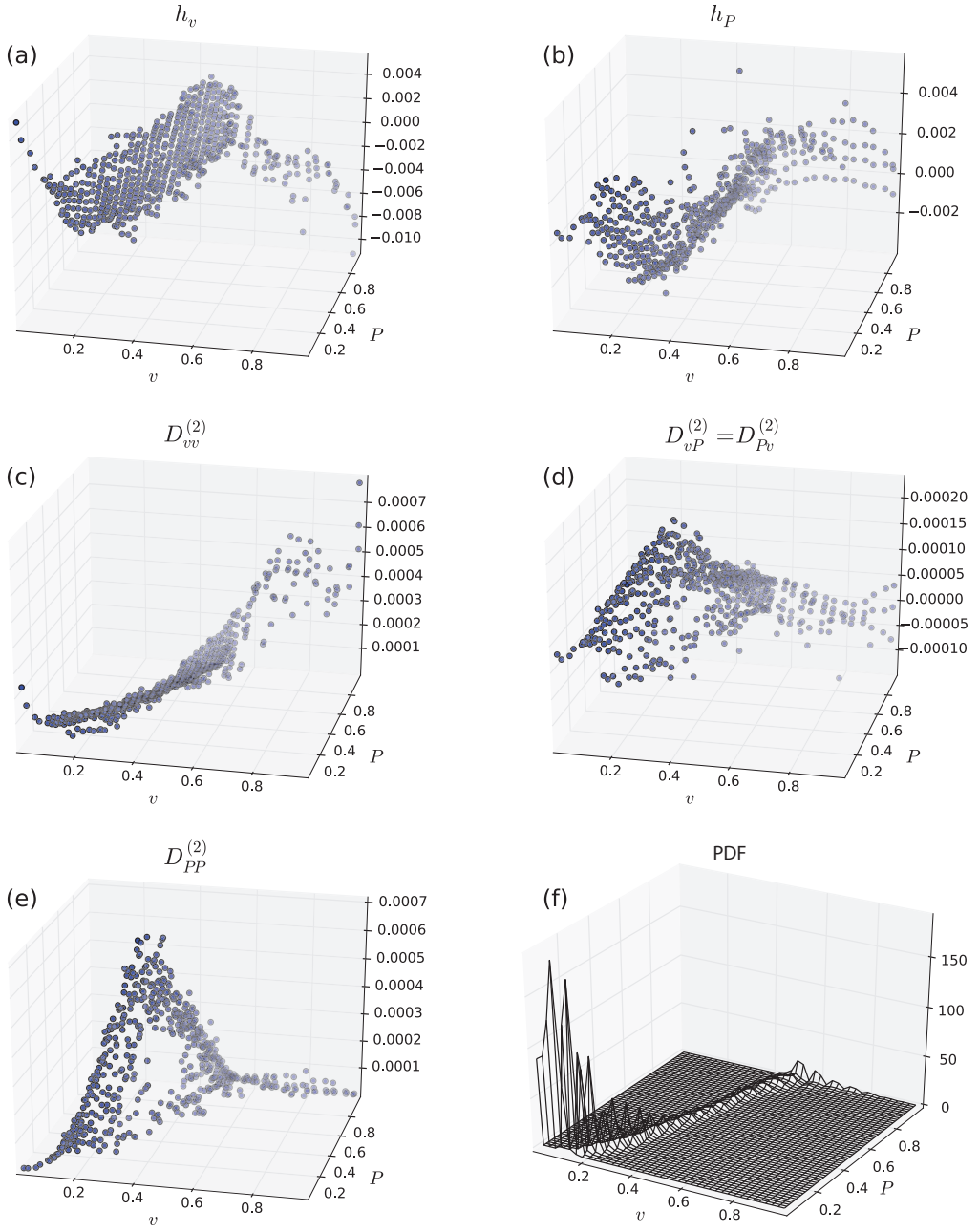


FIG. 4. (Color online) Drift and diffusion coefficients defining the coevolution of P and v: (a) h_v, (b) h_P, (c) $D_{vv}^{(2)}$, (d) $D_{vP}^{(2)} = D_{Pv}^{(2)}$, and (e) $D_{PP}^{(2)}$. The PDF of both variables is shown in (f).

which the turbine was designed and at which it operates at an optimal regime. It can therefore be concluded from our analysis that the turbine has been well selected and it remains to be seen if similar conclusions can be drawn when applying our method to arrays of turbines.

Finally, region III is characterized by frequent low-speed events with a power production below $P \approx 0.6 P_{\max}$, containing another attraction point at $v_{\text{mode}} \approx 0.14 v_{\max}$.

In previous works [5] the drift vector field around the performance curve was parallel to the power production axis. In Fig. 5 the vector field tends to be tilted towards the performance curve, because the data analyzed was sampled with a much smaller frequency, and therefore the time between successive measures is sufficiently large to observe the convergence to the stable fixed points.

Such observations can be more clearly understood by considering Fig. 5 together with the marginal PDF of the wind speed shown in its inset. The distribution of observed values for the wind speed follows approximately a Weibull distribution, as is known from the literature [21]. In our case we observe significant deviations at the tail, which shows a bump. This deviation can be explained with the help of Fig. 2(c) and by recalling that there are correlations between the wind speed with each one of the main wind directions (as highlighted in Sec. II), resulting in the bimodality of the wind speed. The bump in the distribution for the wind

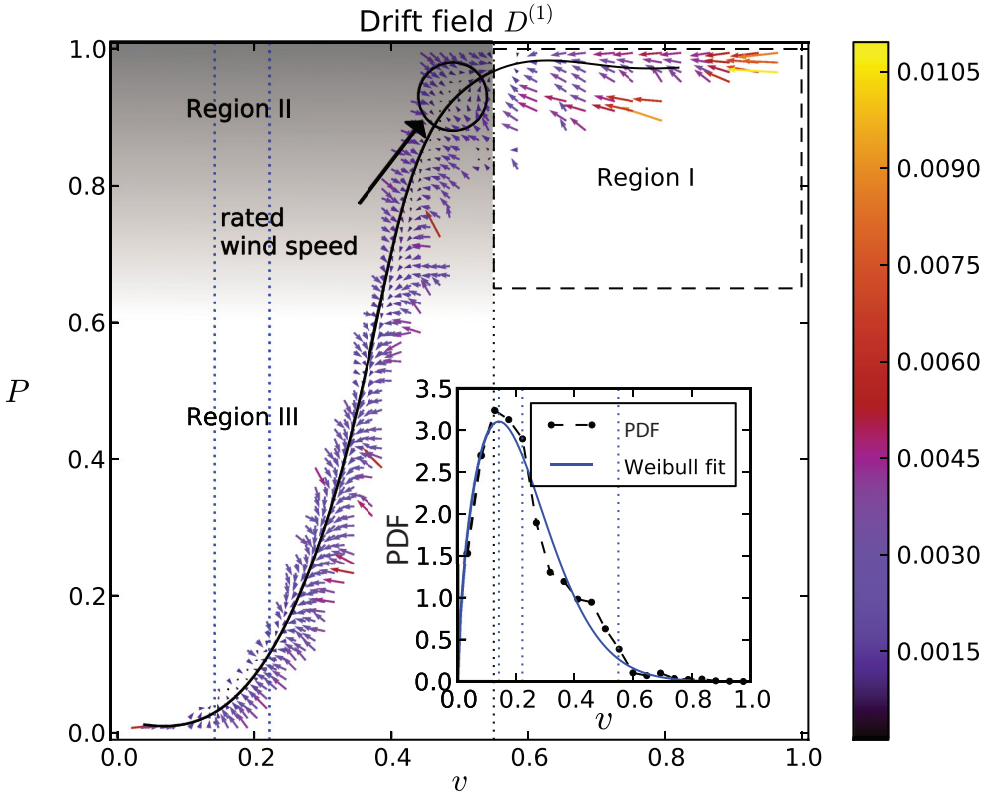


FIG. 5. (Color online) Drift vector $\mathbf{D}^{(1)}(P,v) = [h_P(P,v), h_v(P,v)]$ [see Eq. (5)] in each (P,v) box used to generate the data (see Fig. 3). Three regions can be identified: region I having slow dynamics and regions II and III with fast dynamics. Interestingly, the fixed point spot in region II coincides with the rated speed of the wind turbine. The black curve indicates the performance curve and crosses the bins for which $D^{(1)}(v,P)$ vanishes. In the inset, the marginal probability density function of the wind speed v is well fitted by a Weibull distribution with scale parameter $\lambda \approx 0.25$ and shape parameter $k \approx 1.66$ and a mode $v_{\text{mode}} \approx 0.14 v_{\text{max}}$ [21]. Velocities above $v_{\text{th}} = 0.55 v_{\text{max}}$ are rarely observed, which explain the observed drifts in region I. The gray shading indicates the gradual transition from region II to region III. Dotted vertical lines mark the positions of v_{mode}, $\langle v \rangle$, and v_{th}, respectively.

speed indicates one of the two modes, namely, the one observed at high wind speeds. Therefore, the bimodality of the original data is preserved as bimodality of the Markov chain model.

The two dominant trends identified in region I are compatible with the expected behavior of the power production control system present on the wind turbine. For high values of wind speed, the controller action upon the blade aerodynamics is capable of sustaining the production level despite the expected decrease of the wind speed. One has to consider, however, the time scales involved. With a 10-min resolution of the original data, it is not possible to directly observe the rapid controller action on the blades, only the average behavior of the controller as well as actions that occur on larger time scales, such as the rotation of the tower. However, even at large sampling times, the data set catches some events beyond the power curve and the subsequent conditional moments mirror the controller action that forces the system back on the curve.

Another important application of our method deals with the diffusion matrix. As explained in the previous section, by diagonalizing the diffusion matrix at each point of the phase space one is able to determine the two eigendirections for diffusion. Being orthogonal to each other, these two eigendirections define an ellipse with major and minor axes proportional to the corresponding eigenvalue. Figure 6 shows the diffusion ellipses in phase space. Region I is characterized by the largest ellipses indicating very large fluctuations, while there is an area in region II that presents small fluctuations and corresponds to the fixed point areas identified in the drift field. In the high-slope region of the power curve, the ellipses degenerate, i.e., one eigenvalue is negligible when compared to the other ($\lambda_{\text{min}}/\lambda_{\text{max}} \sim 0$). The inset of Fig. 6 shows in a grayscale the quotient $\lambda_{\text{min}}/\lambda_{\text{max}}$ between the smallest and the largest eigenvalues. White corresponds to zero quotient, while values in $[0.2,1]$ are black. Clearly, a white region indicating a very low ratio of the eigenvalues can be identified, which follows the performance curve shown in Fig. 3. Details concerning our error analysis are described in Appendix B.

One remark is appropriate at this point. As we mention above, the procedure described in this section has been performed on synthetic data generated using a Markov chain model. One might argue that the drift and diffusion coefficients could be extracted directly from the measured data. However, as a process in time, the real data also reflect daily and seasonal variations, which hinders and eventually spoils this approach. It has been found (Sec. IV) that the dynamics of

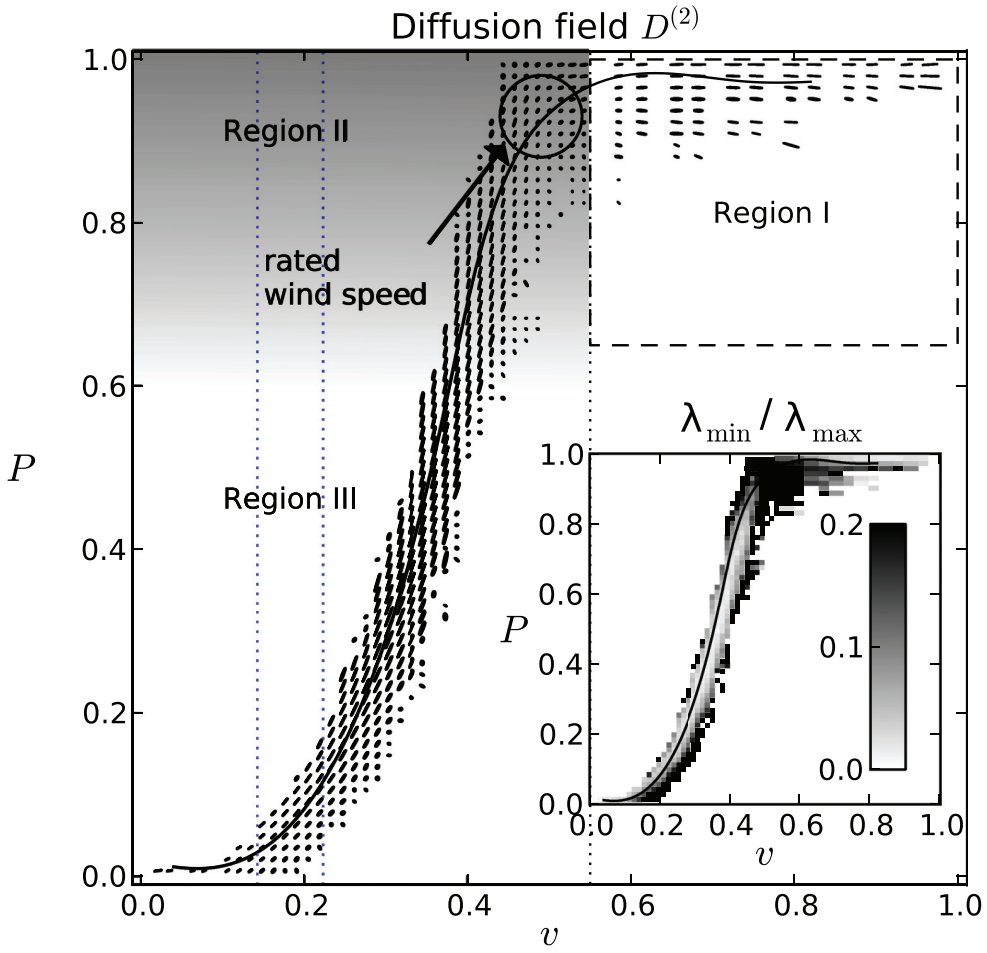


FIG. 6. (Color online) Diffusion ellipses in the power production and wind speed state space. At each box center, the corresponding diffusion ellipse is defined by the two orthogonal eigenvectors of the diffusion matrix $\mathbf{D}^{(2)}$ computed at that center. The principal axes defining the ellipse are aligned along the eigenvectors with a length proportional to the corresponding eigenvalue. Along the performance curve, which gives the functional dependence between both variables P and v, the diffusion ellipses degenerate to a line segment tangential to the curve at each box center. This feature enables one to use the diffusion matrix of any set of variables for deriving their functional relationships. The gray shading indicates the gradual transition from region II to region III. The inset shows the ratio of both eigenvalues $\lambda_{min}/\lambda_{max}$, using a gray scale (0.2 for black and 0 for white). Dotted vertical lines mark the positions of v_{mode}, $\langle v \rangle$, and v_{th}, respectively.

the system are reflected by the presence of two maxima in the joint distribution (one at low v, near v_{mode} and one at high v, near v_{th}). The two-dimensional PDF of the measured data (not shown) closely resembles the one of synthetic data, shown in Fig. 4(f), including the two maxima. Estimating drift and diffusion coefficients directly from the measured data, however, does not allow us to reproduce this distribution (not shown). The MCM has been found to reproduce the relevant statistical and dynamical features of the fluctuations observed in the real data [6]. Given that (i) the MCM produces a transition matrix that maximizes the likelihood of distributions [6], (ii) its errors are known and small (see Appendix B), (iii) it faithfully reproduces the joint PDF, and (iv) the analysis using the MCM yields the correct physical dynamics and fixed points (see Sec. IV), it is reasonable to take the simulated data as the aperiodic process corresponding to the time series of power and wind speed.

V. DERIVING THE PERFORMANCE CURVE FROM UNIVARIATE STOCHASTIC DYNAMICS

The inset of Fig. 6 shows that along the performance curve one eigenvalue is typically much larger than the other. This is an indication that in fact P is a function of v, which in the case of power production and wind speed yields the performance curve drawn in Figs. 3, 5, and 6. To see this one first takes P and v as two general variables fulfilling Eqs. (5) and observes that if $P \equiv P(v)$ there are not two independent stochastic forces, but only one, yielding for v [17] and for P

$$\frac{dv}{dt} = \tilde{h}_v(v) + \tilde{g}_v(v)\Gamma, \tag{6a}$$

$$\frac{dP}{dt} = \tilde{h}_P(P) + \tilde{g}_P(P)\Gamma, \tag{6b}$$

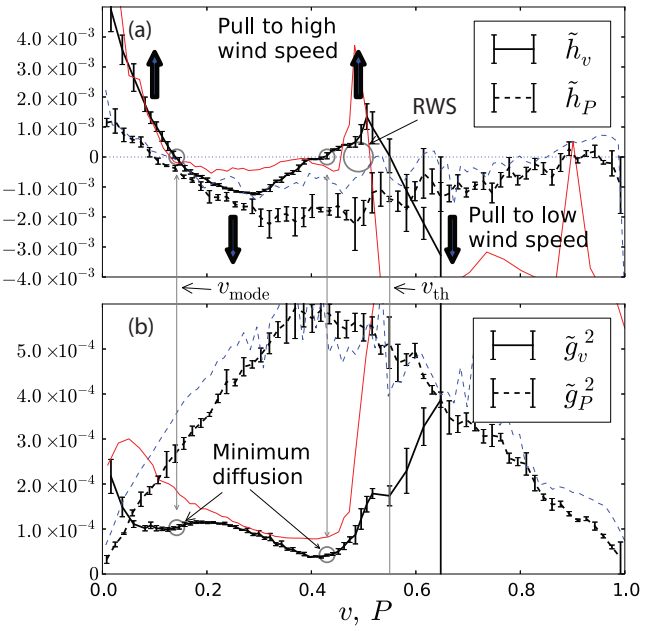


FIG. 7. (Color online) Uncovering properties of wind turbines by analyzing data series of wind speed v and power P separately [see Eqs. (6a) and (6b)]. The horizontal axis indicates the value of $v/v_{\rm max}$ for $\tilde{h}_v$ and $\tilde{g}_v^2$ and the value of $P/P_{\rm max}$ for $\tilde{h}_P$ and $\tilde{g}_P^2$. The large circle denoted by RWS marks the region of the rated wind speed also seen in Figs. 5 and 6. Lines without error bars indicate the same Kramers-Moyal coefficient functions, this time derived directly from the transition matrix, yielding similar results: $\tilde{h}_v$ [red solid line in (a)], $\tilde{h}_P$ [blue dashed line in (a)], $\tilde{g}_v^2$ [red solid line in (b)], and $\tilde{g}_P^2$ [blue dashed line in (b):].

where functions $\tilde{h}$ and $\tilde{g}$ are of course different from the drift and diffusion functions defined above in Eqs. (5), since only one variable is taken into consideration for the stochastic motion equation.

Consequently, the separate analysis of both wind speed and power production enables one to extract valuable insight about the full dynamics and behavior of the wind turbine with the atmospheric wind. In fact, the two-dimensional analysis of the performance curve summarized in Figs. 5 and 6 can indeed be accessed through a one-dimensional stochastic analysis of each variable P and v separately.

Figure 7 shows the drift and diffusion of both the wind speed and power production determined for the model described by Eqs. (6a) and (6b). The drift of the wind speed $\tilde{h}_v$ has three zeros. These zeros correspond to three fixed points, two stable ($v \simeq v_{\rm mode} \simeq 0.14 v_{\rm max}$ and $v \simeq v_{\rm th} \simeq 0.55 v_{\rm max}$) and one unstable at $v_{\rm UFP} \simeq 0.42 v_{\rm max}$. Thus, for wind speed below $v_{\rm UFP}$, the airflow is unstable and unsuited for power production, while wind speeds above $v_{\rm UFP}$ promote power production. The first zero of $\tilde{h}_v$ indicates approximately the mode of the wind speed distribution (compare with the inset in Fig. 5) and the other two zeros mark the transition between two different regions identified above in Figs. 5 and 6. The transition between regions I and II is marked by $v_{\rm th}$. The transition between regions II and III is more subtle and deals with the zero at $v_{\rm UFP}$ and with the functional dependence of the two variables: It is located at the transition region of $0.6 P_{\rm max} \lesssim P \lesssim 0.8 P_{\rm max}$, which corresponds to $v_{\rm UFP} \simeq 0.42 v_{\rm max}$ (cf. Fig. 5). The transition between regions III and II is also located at a minimum of the diffusion $\tilde{g}_v$ for the wind velocity. Moreover, Fig. 7(a) also shows that positive drifts are located at small wind speeds (up to $v_{\rm mode}$) and for region II (see Figs. 5 and 6). At $v_{\rm UFP}$, the drift changes to a positive value. In other words, above $v = v_{\rm UFP}$ the expected change of the wind speed is towards higher values. The drift reaches a maximum on region II prior to a steep change towards negative values. In this region, the wind speed values are not in the range of the extreme weather conditions and are also not as frequent as the lower wind speed values. However, it is frequent enough to be associated with a commonly repeated pattern, i.e., the bimodal pattern of the wind direction, in which the airflow is mainly induced by thermal differences. This pattern is responsible for most of the power production in this wind turbine and the main reason for a second attraction point at $v_{\rm th}$. For higher wind speed values, the drift changes again to negative values (region I) since very high wind speed is usually of short duration, i.e., extreme wind gusts. Moreover, Fig. 7(b) shows where the power stochasticity is maximal, i.e., the region of the highest diffusion values $\tilde{g}_P$, which is the high-slope region of the power curve for $0.3 P_{\rm max} \lesssim P \lesssim 0.6 P_{\rm max}$.

Parallel to our stochastic approach, we also derive coefficients $D^{(1)}$ and $D^{(2)}$ directly from the transition matrix. The results are similar to the ones described above, as one can see in Fig. 7 (thin gray solid and dashed lines), where the deviations from the coefficients derived from our analysis are due to the prescribed binning of phase space, i.e., to the number of states chosen for the transition matrix. Taking all the above observations into account, one concludes that there is strong agreement between the regions defined in the context of Figs. 5 and 6 and the sign of the wind speed drift.

Having analyzed separately both properties v and P, we continue by showing that from the drift and diffusion coefficients $\tilde{h}_v$, $\tilde{g}_v$, $\tilde{h}_P$, and $\tilde{g}_P$, one obtains a functional dependence between power production and wind speed. To that end we assume that Eq. (6a) holds for v and that the other variable P is an exclusive function $P(v)$ of v. Thus we can take the Itô-Taylor expansion [17] of its differential

$$\begin{aligned} dP(v) &= P(v+dv) - P(v) \\ &= \frac{dP}{dv} dv + \frac{1}{2}\frac{d^2P}{dv^2} dv^2 + \mathcal{O}(dv^3) \\ &= \left(\frac{dP}{dv}\tilde{h}_v + \frac{1}{2}\frac{d^2P}{dv^2}\tilde{g}_v^2 \right) dt + \frac{dP}{dv}\tilde{g}_v dw \end{aligned} \tag{7}$$

using the differential $dv = \tilde{h}_v(v)dt + \tilde{g}_v(v)dw$. Therefore, identifying

$$\tilde{h}_P = \frac{dP}{dv}\tilde{h}_v + \frac{1}{2}\frac{d^2P}{dv^2}\tilde{g}_v^2, \tag{8a}$$

$$\tilde{g}_P = \frac{dP}{dv}\tilde{g}_v, \tag{8b}$$

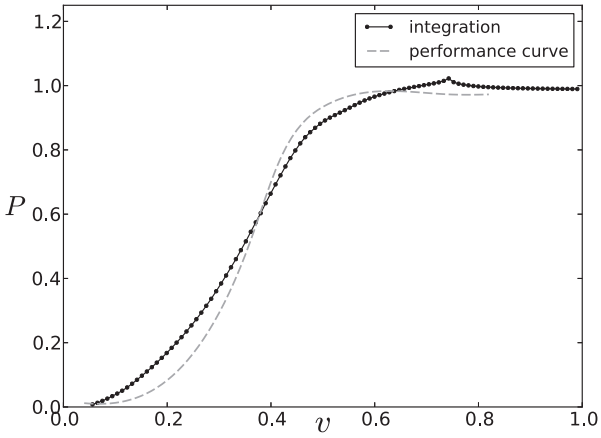


FIG. 8. Power production P as a function of wind speed v by integration [see Eq. (7)]. Both Eqs. (8) are fulfilled, analyzing both series P and v separately [see Eqs. (6a) and (6b)].

which can be solved with respect to the two derivatives of $P(v)$, yields the numerical integration scheme as follows:

$$\begin{aligned} P(v+\Delta t) &= P(v) + \left.\frac{dP}{dv}\right|_{v,P(v)} \Delta v \\ &\quad + \frac{1}{2}\left.\frac{d^2P}{dv^2}\right|_{v,P(v)} (\Delta v)^2 + \mathcal{O}((\Delta v)^3) \\ &= P(v) + \frac{\tilde{g}_P(P(v))}{\tilde{g}_v(v)}\Delta v \\ &\quad + \frac{1}{2}\frac{\tilde{h}_P(P(v))\tilde{g}_v(v) - \tilde{h}_v(v)\tilde{g}_P(P(v))}{(\tilde{g}_v(v)^3}(\Delta v)^2 \\ &\quad + \mathcal{O}((\Delta v)^3). \end{aligned} \tag{9}$$

Figure 8 shows the integration of $dP(v)$ for the condition $P_0(v_0) = 0$ for $v_0 = 0$. The deviations can be attributed to the fact that Eqs. (6a) and (6b) are strictly only valid in the regions where the eigenvalues of the diffusion matrix show a large difference between them, $\lambda_{\min} \ll \lambda_{\max}$ (cf. the insets of Figs. 5 and 6 where $0.2 \leqslant v \leqslant v_{\text{th}}$). In addition, the integration Eq. (9) is valid only on the performance curve and applying Eqs. (6a) and (6b) therefore also neglects the asymmetry of the drift functions with respect to this curve. Both deviations are a natural consequence of having treated two dependent variables P and v as separate stochastic variables.

However, by doing so, two important features can be observed. First, the diffusion matrix $\mathbf{D}^{(2)}(P,v)$ has rank one, i.e., one of its eigenvalues can be neglected in comparison to the other. Second, from the functions h and g in Eqs. (6), it is possible to determine the functional dependence between both variables. More details are given in Appendix C.

VI. DISCUSSION AND CONCLUSIONS

Investigating a wind turbine from a real wind park, we report the reconstruction of the stochastic performance curve of the variables of wind speed and power production, using both drift and diffusion coefficients. These coefficients, describing the respective deterministic and stochastic interactions of wind field, turbine aerodynamics, and controller action, are estimated from a synthetic time series generated using a Markov chain model of the original measurement data. We argue that this reconstruction is superior to a direct evaluation of the measurements.

As a main finding we present the fact that the reconstruction of the power curve using both drift and diffusion coefficients uncovers additional information not visible in an analysis of the drift field alone [10], even though we are using measured data of a very low measurement rate as model input. Specifically, our analysis reveals the existence of various distinct regions in the wind speed-power production plane. In addition, we have been able to reconstruct the power curve from the drift and diffusion coefficients using a method that should be able to uncover functional relationships between stochastic variables in a wide range of experimental setups. It should be noted that it is possible to infer the correlation between v and P from the joint PDF in Fig. 4(f) alone; however, such an approach neglects the dynamical behavior of the system.

Without our approach summarized in Figs. 5–7 one could not so surely claim the existence of three separate regions. Moreover, attached to these three regions we detected three fixed points of the dynamics, two of them stable and one unstable. These three fixed points are only shown clearly in Fig. 7, after performing the one-dimensional analysis with our method. In particular, considering the upper stable fixed point, while its coincidence with peak production can be identified directly in the (P,v) joint PDF, the analysis sketched in Fig. 7 allows us to detect the region that belongs to its basin of attraction, within which drift drives the trajectories to the upper stable fixed point, whereas beyond this region trajectories are pushed to the lower stable fixed point.

Finally, information about how trajectories diffuse in phase space, i.e., the entire dynamics of the system, can only be obtained completely after extracting the drift and diffusion fields together with the fixed points of the drift field and the principal directions of diffusion. Only after analyzing the separation of diffusion eigenvalues visible in Fig. 6 can we postulate the existence of a single diffusive force underlying Eqs. (6a)–(C5).

Although there are limitations in increasing the extracted power of a wind turbine, typically described by the power coefficient that has a maximum value given by the Betz limit [22], our approach may be helpful in obtaining a better understanding of the complex dynamics that determines power production in wind turbines. Indeed, we believe that such a stochastic description, if applied to an entire wind park, would enable one to better quantify the risk associated with the estimate of global energy production. The global energy production of a wind park is usually determined by financial constraints, i.e., by decision making of how much energy one must buy or sell in the market to compensate for the energy production fluctuations. Therefore, we are now extending this methodology in order to consider coupled systems of wind turbines in nearby locations, using direct measures of power and wind speed as well as simulated data.

ACKNOWLEDGMENTS

The authors thank Matthias Wächter, David Kleinhans, and Maria Haase for useful discussions and GENERG, SA for providing the original data. The authors

acknowledge partial support under Grants No. PEst-OE/FIS/UI0618/2011 and No. FCOMP-01-0124-FEDER-016080. Also, F.R. (Grant No. SFRH/BPD/65427/2009), T.S. (Grant No. SFRH/BD/86934/2012), and P.G.L. (Grant No. Ciência 2007) thank Fundação para a Ciência e a Tecnologia (FCT) for financial support. This work was part of a bilateral cooperation DRI/DAAD/1208/2013 supported by FCT and Deutscher Akademischer Auslandsdienst.

APPENDIX A: ESTIMATION OF THE MARKOV CHAIN TRANSITION MATRIX

The Markov chain transition probabilities were not obtained by employing the usual maximum likelihood (ML) estimator, but by using a modified likelihood function that combines information from one-step and two-step transitions, as described in this appendix. This modified ML yields lower variance estimates for the transition probabilities (see Appendix B).

Let $\{s_i | i \in \{0,\ldots,n\}\}$, $n \in \mathbb{N}$, be the state space of the Markov chain model, where each state is a combination of a wind power, speed, and direction state. Thus the wind power, speed, and direction time series can be transferred into a stream s of states, i.e., $\mathbf{s} = \{s_{i_0}, s_{i_1}, s_{i_2}, \ldots, s_{i_{m-2}}, s_{i_{m-1}}, s_{i_m}\}$, where $m \in \mathbb{N}$ denotes the length of the time series and s_{i_k} denotes the state the Markov process assumes at time k with $i_k \in \{1,\ldots,n\}$ $\forall k \in \{0,\ldots,m\}$. Let $p_{i,j}$ denote the probability of the process to move from state s_i, at time k, to state s_j, at time $k+1$. Then the likelihood function $\mathcal{L}^{(1)}$, i.e., the probability of the observed series of states $\mathbf{s}$, given the transition probabilities $p_{i,j}$ for one-step transitions, is

$$\mathcal{L}^{(1)} = P(s_{i_0})\, p_{i_0,i_1} p_{i_1,i_2} \cdots p_{i_{m-2},i_{m-1}} p_{i_{m-1},i_m}. \tag{A1}$$

Since the probability of the process being in state s_{i_0} is constant, the maximum likelihood estimator can thus be written as the maximum of $\mathcal{L}^{(1)} = \prod_{(i,j)\in\mathcal{S}_1} p_{i,j}$ subjected to $p_{i,j} \geqslant 0$ and $\sum_{j=0}^{n} p_{i,j} = 1$, with $i,j = 1,\ldots,n$, where $\mathcal{S}_1$ is the set of all one-step transitions (i,j) observed in the state stream $\mathbf{s}$.

Solving the above optimization problem is equivalent to minimizing the negative log-likelihood function $\mathcal{L}^{(1)}_{\log} = -\sum_{(i,j)\in\mathcal{S}_1} \log p_{i,j}$, subjected to the same constraints. For practical purposes we consider henceforth the log-likelihood function $\mathcal{L}^{(1)}_{\log}$.

Next we consider only two-step transitions. Taking $p'_{i,j}$ as the probability of the process moving from state s_i, at time k, to state s_j, at time $k+2$, the maximum likelihood estimators are given by the maximum of $\mathcal{L}^{(2)} = \prod_{(i,j)\in\mathcal{S}_2} p'_{i,j} = \prod_{(i,j)\in\mathcal{S}_2} \sum_{k=1}^{n} p_{i,k} p_{k,j}$, with $\mathcal{S}_2 = \mathcal{S}_2^{\text{odd}} \cup \mathcal{S}_2^{\text{even}}$, where $\mathcal{S}_2^{\text{odd}} = \{(i_1,i_3),(i_3,i_5),\ldots,(i_{m-3},i_{m-1})\}$ (odd time points) and $\mathcal{S}_2^{\text{even}} = \{(i_0,i_2),(i_2,i_4),\ldots,(i_{m-2},i_m)\}$ (even time points), assuming without loss of generality that m is even. Here we use the Chapman-Kolmogorov equations to express the two-step transitions $p'_{i,j}$ in terms of one-step transitions $p_{i,j}$ with the same constraints as above. The corresponding log-likelihood function, whose minimization defines an equivalent optimization problem, is $\mathcal{L}^{(2)}_{\log} = -\sum_{(i,j)\in\mathcal{S}_2} \log(\sum_{k=1}^{n} p_{i,k} p_{k,j})$.

Finally, as described in Ref. [6], we minimize the objective function given by the sum of the two log-likelihood functions above, namely,

$$\mathcal{L}_{\log} = \mathcal{L}^{(1)}_{\log} + \mathcal{L}^{(2)}_{\log}, \tag{A2}$$

and under the same constraints. Higher-order Markov models were also tested, but did not show considerable improvement compared to the estimator $\mathcal{L}_{\log}$ in Eq. (A2). Notice that the minimization of $\mathcal{L}_{\log}$ in Eq. (A2) for the series of states $\mathbf{s}$ is equivalent to minimizing the one-step functional $\mathcal{L}^{(1)}_{\log}$ of the augmented state stream $\mathbf{s}' = s_{i_0}, s_{i_1}, s_{i_2}, \ldots, s_{i_{m-2}}, s_{i_{m-1}}, s_{i_m}, S, s_{i_1}, S, s_{i_3}, S, s_{i_5}, \ldots, s_{i_{m-3}}, S, s_{i_{m-1}}, S, s_{i_0}, S, s_{i_2}, S, s_{i_4}, \ldots, s_{i_{m-2}}, S, s_{i_m}$, where S stands for voids in the state stream.

Concerning the state space partition, we tried several different partitions of the power-speed-direction space. A comparison of the histograms of the original data and the data synthesized using these models showed that the model with the highest resolution of the binning in the power-speed plane is $80 \times 60 \times 12$, which reproduces the original data most accurately.

APPENDIX B: ERROR ANALYSIS

A lower bound for the uncertainty of the transition probabilities is provided by the Cramer-Rao bound (CRB) since there is no simple expression to determine the parameters variance for the two-step estimator, described in Sec. VI. The CRB for maximum likelihood estimators with constraints on the parameter space can be computed by using the following equation [23]:

$$\text{CRB} = U^T (U H U^T)^{-1} U, \tag{B1}$$

where U is an orthonormal matrix spanning the null space of the Jacobian J of the parameter's equality constraints and H is the Hessian of the objective function, defined in Appendix A.

The uncertainty is linked to the number of observed transitions in the data. Results show that a high value for the CRB ($\sigma_{p_{ij}} > 0.1$) is always associated with infrequent transitions (fewer than four observations in the 2-yr data set). For transition probabilities p_{ij} with more than 25 observations, 90% of the $\sigma_{p_{ij}}$ values are smaller than $0.25\, p_{ij}$.

To assess the performance of the two-step estimator, its CRB is compared with the CRB of the one-step estimator, applied to the same 2-yr data set. Results shows that for the transition probabilities with a high CRB value, the two-step estimator provides a lower bound. In the remaining cases, the difference is not significant.

Another source of errors stems from our use of the direct estimation of the Kramers-Moyal (KM) coefficients from the synthetic time series. This method has been found to introduce three principal types of errors [18,24,25].

The first error accounts for the statistical variation of counts N in each bin used for the calculation of the conditional moments. It can be shown [25] that this error decreases with $1/\sqrt{N}$. Since we can generate synthetic time series of arbitrary length and therefore arbitrarily high N in each bin under consideration, we can neglect this error.

Second, estimation of the KM coefficients uses an expansion of the Fokker-Planck operator in powers of the temporal increment τ, neglecting higher orders in τ. It is known [24] that this finite-time expansion induces an erroneous count $M_E^{(i)}$ of the ith conditional moment $M^{(i)}$, namely, $M_{1,E} = \tau D^{(1)} +$

$\frac{\tau^2}{2}(D^{(1)}D_x^{(1)} + D^{(2)}D_{xx}^{(1)})$ and $M_{2,E} = 2\tau D^{(2)} + \tau^2(D^{(1)}D^{(1)} + 2D^{(2)}D_x^{(1)} + D^{(1)}D_x^{(2)} + D^{(2)}D_{xx}^{(2)})$. We have calculated these errors numerically, using the estimated KM coefficients $D^{(i)}$ and their numerical first and second derivatives $D_x^{(i)}$ and $D_{xx}^{(i)}$, and found them to be generally within a few percent of the estimated KM coefficients.

A third source of error is the finite size of the bins used for the calculation of the conditional moments. Again, it can be shown [25] that this finiteness induces an erroneous count of the moments (and corresponding coefficients)

$$D_E^{(k)}(x_0) = \frac{\int_{x_0-\Delta x}^{x_0+\Delta x} D^{(k)}(x_0)p(x)dx}{\int_{x_0-\Delta x}^{x_0+\Delta x} p(x)dx}, \tag{B2}$$

where x_0 and Δx are the respective bin centers and bin widths and $p(x)$ is the stationary distribution of the stochastic variable. We numerically investigated this error using fits for both the stationary distributions and KM coefficients inside the bins and found it to be generally in the few-percent regime too, increasing considerably only at the edges of the regions investigated, where the KM functions become steeper.

The effect of the second and third error sources is indicated by error bars in Fig. 7.

APPENDIX C: STOCHASTIC ANALYSIS OF TWO DEPENDENT VARIABLES

In general, whenever the diffusion matrix $\mathbf{D}^{(2)}(P,v)$ has rank one, the set of variables (P,v) has in fact only one independent stochastic source and therefore Eqs. (5) reduce to Eqs. (6), where the stochastic force Γ is the same for both variables. Consequently, one can write the differential of v as

$$dv = \tilde{h}_v dt + \tilde{g}_v dW, \tag{C1}$$

where the stochastic differential dW is the *same* as the one in Eq. (7). Since both variables P and v are driven by the same stochastic forces, one can take only one of them as the stochastic variable, say, v, and the other one as a function of v and t alone, $P \equiv P(v,t)$. In that way one incorporates all stochastic contributions into v. Mathematically this implies that we can write the differential of P as

$$dP = \frac{\partial P}{\partial v}dv + \frac{\partial P}{\partial t}dt, \tag{C2}$$

where both partial derivatives are derived from the functions $\tilde{h}$ and $\tilde{g}$ alone [see Eq. (6)]. Indeed, rewriting Eq. (7) as

$$dP(v) = \frac{dP}{dv}(\tilde{h}_v dt + \tilde{g}_v dW) + \frac{1}{2}\frac{d^2P}{dv^2}\tilde{g}_v^2 dt \tag{C3}$$

and using Eq. (C1) together with the relations in Eq. (8) yields

$$dP = \frac{\tilde{g}_P}{\tilde{g}_v}dv + \left(\tilde{h}_p - \tilde{h}_v\frac{\tilde{g}_P}{\tilde{g}_v}\right)dt. \tag{C4}$$

This equation means that P is a function of t and v, which contains all stochastic contributions. Consequently, the partial derivatives in Eq. (C2) are

$$\frac{\partial P}{\partial v} = \frac{\tilde{g}_P}{\tilde{g}_v}, \tag{C5a}$$

$$\frac{\partial P}{\partial t} = \tilde{h}_p - \tilde{h}_v\frac{\tilde{g}_P}{\tilde{g}_v}. \tag{C5b}$$

The first partial derivative shows that the variation of the production power by speed variations equals the quotient of the corresponding diffusion amplitudes in time. The second partial derivative describes the local power production, which is given by the power production drift $\tilde{h}_p$, after subtracting the contribution of the wind speed drift $\tilde{h}_v$ in the variation of P due to v.

[1] W. M. Adams, The Future of Sustainability: Re-thinking Environment and Development in the Twenty-first Century: Report of the IUCN Renowned Thinkers Meeting (2006), http://cmsdata.iucn.org/downloads/iucn_future_of_sustanability.pdf.

[2] U. S. Department of Energy Report 20% Wind Energy by 2030, 2010 (unpublished), available at http://www.windpoweringamerica.gov.

[3] A. Rauh and J. Peinke, J. Wind Eng. Ind. Aerodyn. **92**, 159 (2004).

[4] P. Milan, M. Wächter, and J. Peinke, Phys. Rev. Lett. **110**, 138701 (2013).

[5] J. Gottschall and J. Peinke, J. Phys. Conf. Ser. **75**, 012045 (2007).

[6] V. V. Lopes, T. Scholz, A. Estanqueiro, and A. Q. Novais, in *Proceedings of the 11th International Conference on Environment and Electrical Engineering (EEEIC), 18–25 May 2012* (IEEE, Piscataway, NJ, 2012), pp. 770–775.

[7] A. D. Sahin and Z. Sen, J. Wind Eng. Ind. Aerodyn. **89**, 263 (2001).

[8] V. V. Vasconcelos, F. Raischel, M. Haase, J. Peinke, M. Wächter, P. G. Lind, and D. Kleinhans, Phys. Rev. E **84**, 031103 (2011).

[9] F. Raischel, A. Russo, M. Haase, D. Kleinhans, and P. G. Lind, Phys. Lett. A **376**, 2081 (2012).

[10] E. Anahua, S. Barth, and J. Peinke, Wind Energy **11**, 219 (2008).

[11] M. Wächter, H. Heißelmann, M. Hölling, A. Morales, P. Milan, T. Mücke, J. Peinke, N. Reinke, and P. Rinn, J. Turbul. **13**, N26 (2012).

[12] D. Kleinhans, Phys. Rev. E **85**, 026705 (2012).

[13] S. J. Lade, Phys. Rev. E **80**, 031137 (2009).

[14] H. Risken, *The Fokker-Planck Equation* (Springer, Heidelberg, 1984).

[15] J. Gradišek, R. Friedrich, E. Govekar, and I. Grabec, Meccanica **38**, 33 (2003).

[16] F. O. Hocaoglu, O. N. Gerek, and M. Kurban, in *Proceedings of the 10th International Conference on Probabilistic Methods Applied to Power Systems, 25–29 May 2008, PMAPS '08* (IEEE, Piscataway, NJ, 2008), pp. 1–4.

[17] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 1997).

[18] R. Friedrich, J. Peinke, M. Sahimi, and M. R. R. Tabar, Phys. Rep. **506**, 87 (2011).

[19] P. G. Lind, M. Haase, F. Böttcher, J. Peinke, D. Kleinhans, and R. Friedrich, Phys. Rev. E **81**, 041125 (2010).
[20] A. M. van Mourik, A. Daffertshofer, and P. J. Beek, Biol. Cybern. **94**, 233 (2006).
[21] J. A. Carta, P. Ramírez, and S. Velázquez, Renew. Sust. Energ. Rev. **13**, 933 (2009).
[22] T. Burton, D. Sharpe, N. Jenkins, and E. Bossanyi, *Wind Energy Handbook* (Wiley, London, 2001).
[23] P. Stoica and B. C. Ng, IEEE Signal Process. Lett. **5**, 177 (1998).
[24] R. Friedrich, C. Renner, M. Siefert, and J. Peinke, Phys. Rev. Lett. **89**, 149401 (2002).
[25] D. Kleinhans, Master thesis, German University System, 2005.

Chapter 5

A cyclic time dependent Markov process to model daily patterns in wind turbine power production

The two previous chapters described the analysis of wind energy production in the Langevin frameworks as a two stage process. In the first stage, the empirical measurement data were modeled with a Markov chain, in the second stage, data synthesized from this Markov chain were analysed using the Langevin method.
As it stands, this method is only capable of extracting information concerning long-term behavior from the data. In wind energy production however, certain patterns repeat themselves on a daily basis due to the solar cycle. This paper describes a way to incorporate these patterns into the modeling stage, by extending the model to a time-dependent Markov process. A challenge that arises with this extension is the sheer volume of new parameters, since every model parameter now needs to be estimated for each available time of the day. In the case of the industry standard of 10-minute average data this yields 144 times per day for which the transition matrix needs to be computed. The paper presents a way to reduce the amount of parameters by modeling the time-dependent transition probabilities in terms of Bernstein polynomials, effectively decreasing the amount of parameters to the degree of the approximating polynomial.
The paper describes the estimation procedure based on a convex constrained optimization problem and shows the agreement of the model with the data. Moreover, the use of the model aside from data synthesis is demonstrated by anaysing wind power production persistence for different times of the day.

Energy 67 (2014) 557–568

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

A cyclic time-dependent Markov process to model daily patterns in wind turbine power production

Teresa Scholz, Vitor V. Lopes*, Ana Estanqueiro

LNEG, National Laboratory for Energy and Geology, Estrada do Paço do Lumiar, 22, 1649-038 Lisboa, Portugal

ARTICLE INFO

Article history:
Received 18 September 2013
Accepted 30 December 2013
Available online 30 January 2014

Keywords:
Cyclic Markov process
Wind power
Persistence
Diurnal pattern

ABSTRACT

Wind energy is becoming a top contributor to the renewable energy mix, which raises potential reliability issues for the grid due to the fluctuating nature of its source. To achieve adequate reserve commitment and to promote market participation, it is necessary to provide models that can capture daily patterns in wind power production. This paper presents a cyclic inhomogeneous Markov process, which is based on a three-dimensional state-space (wind power, speed and direction). Each time-dependent transition probability is expressed as a Bernstein polynomial. The model parameters are estimated by solving a constrained optimization problem: The objective function combines two maximum likelihood estimators, one to ensure that the Markov process long-term behavior reproduces the data accurately and another to capture daily fluctuations. A convex formulation for the overall optimization problem is presented and its applicability demonstrated through the analysis of a case-study. The proposed model is capable of reproducing the diurnal patterns of a three-year dataset collected from a wind turbine located in a mountainous region in Portugal. In addition, it is shown how to compute persistence statistics directly from the Markov process transition matrices. Based on the case-study, the power production persistence through the daily cycle is analyzed and discussed.

1. Introduction

The EC European Parliament objective to achieve 20% of the consumed energy from the renewable energy sector by 2020 introduced a serious challenge to the planning and operating of power systems. Wind energy is becoming a top contributor to the renewable energy mix due to rather high capacities and generation costs that are becoming competitive with conventional energy sources [28]. However, wind energy systems suffer from a major drawback, the fluctuating nature of their source, which affects the grid security, the power system operation and market economics. There are several tools to deal with these issues, such as the knowledge of wind power persistence and wind speed or power simulation. Persistence is related to stability properties and can provide useful information for bidding on the electricity market or to maintain reliability, e.g. by setting reserve capacity.

Wind power or speed simulation can be used to study the impact of wind generation on the power system. For this task, a sufficiently long time series of the power output from the wind plants should be used. However, real data records are commonly of short length and thus synthetic time series are generated by stochastic simulation techniques to model wind activity [16]. Shamshad et al. [23] used first and second-order Markov chain models for the generation of hourly wind speed time series. They found that a model with 12 wind speed states (1 m/s size) can capture the shape of the probability density function and preserve the properties of the observed time series. Additionally, they concluded that a second-order Markov chain produces better results. Nfaoui et al. [15] compared the limiting behavior of their Markov chain model with the data histograms gotten from hourly averaged wind speed and showed that the statistical characteristics were faithfully reproduced. Sahin and Sen [22] reported the use of a first-order Markov chain approach to simulate the wind speed, where: a) both transitions between consecutive times and within state wind speeds are sampled using a uniform distribution; and, b) extreme states are sampled with an exponential distribution. They showed that statistical parameters were preserved to a significant extent; however, second-order Markov chain models could yield improved results.

Although wind power can be computed from synthetic wind speed time series, Papaefthymiou and Klöckl [16] show that a stochastic model using wind power leads to a reduced number of states and a lower Markov chain model order. They compared a

* Corresponding author.
E-mail addresses: teresa.scholz@lneg.pt (T. Scholz), vitor.lopes@lneg.pt (V.V. Lopes), ana.estanqueiro@lneg.pt (A. Estanqueiro).

0360-5442/$ – see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.energy.2013.12.071

Nomenclature

α_0	initial state distribution at time step $t = 0$
$\beta_\mu^{i,j}$	coefficients of the Bernstein polynomial modeling the transition probability $p_{i,j}(t)$
$1_{\mathcal{A}}$	unit column vector of the same size as subset $\mathcal{A}$
$\mathbf{P}$	$P_0 \bullet \ldots \bullet P_{T-1}$
$\mathcal{A}$	subset of the state space, containing the states of interest for persistence
$\mathcal{S}$	set of observed state transitions
$\mathcal{S}_z$	set of transitions observed in the data together with the scaled time of the day z at which they are observed
ω	weight of the extra transitions added to the objective function
π	stationary distribution of a time-invariant Markov chain
π^*	$\lim_{t\to\infty} \mathbf{P}^t$
π_r	stationary distribution at time r of a time-variant cyclic Markov process
$\pi_r(j)$	stationary probability, of state j at time of the day r
$\pi_r(\mathcal{A})$	vector whose elements are the stationary probabilities of the states in the set $\mathcal{A}$ at time of the day r
τ	persistence
τ_r	time-dependent persistence in a cyclic Markov process
$b_{\mu,k}(z)$	μ-th Bernstein basis polynomial of order k
$E[\,]$	expected value operator
P_t	t-th step transition matrix of a Markov process
$p_{i,j}(t)$	t-th step transition probability of a Markov process
$p_{i,j}^{\text{avg}}$	daily average probability of transition from state s_i to s_j
r_t	remainder of time step t modulo T
S	Markov process state space
s_i	i-th state of a Markov process
T	period of a cyclic Markov process
t	time step of a Markov process
X_t	Markov process
z	scaled time of the day
$\pi_{\mathcal{A}}$	stationary probability distribution of the states in subset $\mathcal{A}$
r	time of the day

Markov chain based method for the direct generation of wind power time series with the transformed generated wind speed. Both the autocorrelation and the probability density function of the simulated data showed a good fit. Thus, they concluded that it is better to generate wind power time series. Chen et al. [7] also modeled wind power by using different discrete Markov chain models: the basic Markov model; the Bayesian Markov model, which considers the transition matrix uncertainty; and, the birth-and-death Markov model, which only allows state transitions between immediately adjacent states. After comparing the wind power autocorrelation function, the authors find the Bayesian Markov model best. Lopes et al. [13] proposed a Markov chain model using states that combine information about wind speed, direction and power. From the transition matrix, they compute statistics, such as the stationary power distribution and persistence of power production, which show a close agreement with their empirical analogs. The model was then used for the two-dimensional stochastic modeling of wind dynamics by Raischel et al. [21]. They aim at studying the interactions between wind velocity, turbine aerodynamics and controller action using a system of coupled stochastic equations describing the co-evolution of wind power and speed. They showed that both the deterministic and stochastic terms of the equations can be extracted directly from the Markov chain model.

The knowledge of wind power production persistence provides useful information to run a wind park and to bid on the electricity market, since it provides information about the expected power steadiness. It can be seen as the average time that a system remains in a given state or a subset of states. Existent literature focuses mainly on wind speed persistence, which is used for assessing the wind power potential of a region. Persistence can be determined directly from the data [20,19]; however, the presence of missing data leads to an underestimate of actual persistence. Alternative methods are based on wind speed duration curves [14,10], the autocorrelation function or conditional probabilities. Koçak [11] and Cancino-Solórzano et al. [5] compare these techniques, and both conclude that wind speed duration curve yields the best results, i.e. results that follow the geographical and climatic conditions of the analyzed sites. Moreover, Cancino-Solórzano et al. [5] analyze the concept of "useful persistence", which is the time schedule series where the wind speed is between the turbine cut-in and cut-out speed. The results gotten from this analysis coincide with the persistence classification obtained using the speed duration curves. In addition, Koçak [12] suggests a detrended fluctuation analysis to detect long-term correlations and analyze the persistence properties of wind speed records. Sigl et al. [24]; Corotis et al. [8] and Poje [19] proposed an approach based on the use of a power law or exponential probability distributions for the persistence of wind speed above and below a reference value. A Markov chain based method to derive the distribution of persistence is introduced by Anastasiou and Tsekos [1], who show its capability on wind speed data.

Most methods in literature of wind speed and power synthesis fail to represent diurnal patterns in the artificial data. However, these are relevant for energy system modeling and design, since their knowledge allows to plan and schedule better. For instance, a power production behavior that best matches demand needs smaller reserve capacity. Recently, Suomalainen et al. [26,25] introduced a method for synthetic generation of wind speed scenarios that include daily wind patterns by sampling a probability distribution matrix based on five selected daily patterns and the mean speed of each day. Carapellucci and Giordano [6] adopt a physical–statistical approach to synthesize wind speed data and evaluate the influence of the diurnal wind speed profile on the cross-correlation between produced energy and electrical loads. The parameters of their model, such as diurnal pattern strength or peak hour of wind speed are determined through a multi-objective optimization, carried out using a genetic algorithm.

This paper introduces a cyclic time-variant Markov model of wind power, speed and direction designed to consider the daily patterns observed in the data. The model can be used to synthesize data for the three variables and is capable of reproducing the daily patterns. Moreover, it allows to compute persistence statistics depending on the time of the day. The paper is organized as follows: Section 2 introduces the proposed model as an extension of the "regular" Markov chain model, which is then used for comparison. Furthermore it is shown, how to compute the time of the day dependent persistence statistics directly from the Markov model transition matrices. In Section 3 the constrained convex optimization problem to get the model parameters is introduced and explained. It is applied to the analysis of a case-study based on a real dataset, Section 4. Since the model describes the joint

statistics for wind power, speed and direction, Section 5 explains how to create synthetic time series for these variables. Section 6 compares the synthesized data of both the time-variant and the time-invariant versions of the model. Moreover, it is shown how the persistence of power production varies through the daily cycle.

2. Time-inhomogeneous Markov model

2.1. Definition

A discrete finite Markov process $\{X_t \in S,\ t \geq 0\}$ is a stochastic process on a discrete finite state space $S = \{s_0, s_1, \ldots, s_n\}$, $n \in \mathbb{N}$, whose future evolution depends only on its current state [9]. This Markov property is expressed mathematically by

$$Pr\{X_{t+1} = s_j | X_t = s_i \wedge X_l \in S\ \forall l = 0, \ldots, t-1\} = Pr\{X_{t+1} = s_j | X_t = s_i\}.$$

$Pr\{X_{t+1} = s_j | X_t = s_i\}$ describes the probability of the Markov process moving to state s_j at time step $t+1$ given that it is in state s_i at time step t and is called the t-th step transition probability, denoted as $p_{i,j}(t)$. Thus, for each time step t the Markov process has an associated transition probability matrix P_t, a n by n matrix with entries $[P_t]_{i,j} = p_{i,j}(t)$ for all $i, j \in \{0, \ldots, n\}$. Each P_t satisfies the following properties: $p_{i,j}(t) \geq 0$ and $\sum_j p_{i,j}(t) = 1\ \forall i, j \in \{0, \ldots, n\}, \forall t$. A Markov process is called cyclic with period $T \in \mathbb{N}$, if T is the smallest number, such that $p_{i,j}(mT + r) = p_{i,j}(r)$ for all m in $\mathbb{N}$, $0 \leq r < T$ [18]. Thus, a cyclic Markov process is described by T transition matrices P_r, $r = 0, \ldots, T-1$. The remainder of time step t modulo T will be denoted as r_t and thus $r_t = r_{t+mT}\ \forall t$, $m \in \mathbb{N}$.

If the transition probabilities are time-independent, i.e. $p_{i,j}(t) = p_{i,j}$, the process is called a (time-homogeneous) Markov chain and its probability matrix $P \in \mathbb{R}^{n+1 \times n+1}$ is given by $[P]_{i,j} = p_{i,j}$. By analogy to the time-dependent case it holds that $p_{i,j} \geq 0$ and $\sum_j p_{i,j} = 1\ \forall i, j \in \{0, \ldots, n\}$.

2.2. Communication classes and irreducibility

2.2.1. Time-invariant Markov chain

The probability of reaching a state s_j from a state s_i in l time steps is given by $P^l(i, j)$, i.e. the l-th power of the transition matrix P. If a state s_j can be reached from a state s_i in a finite number of time steps and vice versa, i.e. $\exists l \in \mathbb{N}\ P^l(i, j) \geq 0 \wedge P^l(j, i) \geq 0$, the states s_i and s_j communicate. All states that communicate with each other are said to be in the same communication class. If all states of a state space are in the same communication class, i.e. if it is possible to reach every state from any other state in a finite number of time steps, the corresponding transition matrix P is called irreducible.

2.2.2. Cyclic time-variant Markov process

A cyclic Markov process with period T is described by T transition matrices P_r, one for each time of the day $r = 0, \ldots, T-1$. The probability of the process reaching state s_j from state s_i in l time steps at time $t = 0$ is given as $(P_0 \cdot \ldots \cdot P_{T-1})^m \cdot P_0 \cdot \ldots \cdot P_{r_l}(i, j)$ with $l = mT + r_l$. For an arbitrary time step t, the formula must be multiplied from the left with the term $P_{rt} \cdot \ldots \cdot P_{T-1}$. Thus, the Markov process is irreducible, if the matrix $\mathbf{P} = P_0 \cdot \ldots \cdot P_{T-1}$ is irreducible, i.e. if $\exists l \in \mathbb{N}$: $\mathbf{P}^l(i, j) \geq 0\ \forall i, j$.

2.3. The stationary distribution

If a Markov chain is irreducible and aperiodic then the long-term statistics are described by the stationary probability distribution $\pi = \lim_{t \to \infty} \alpha_0 P^t$. The distribution is independent of the initial distribution α_0 and satisfies the balance equation $\pi = \pi P$. By the Perron–Frobenius theorem it can be computed as the normalized eigenvector corresponding to the eigenvalue 1 of the transition matrix [17].

In the case of the cyclic time-inhomogeneous Markov process there is also a stationary distribution π_r, for all $r < T$. It can be interpreted as the limiting distribution of the Markov process considering only the datapoints sampled at time of the day r. If the matrix $\mathbf{P}$ is irreducible, i.e. if $\exists \pi^*$, such that $\pi^* = \lim_{t \to \infty} \alpha_0 \cdot \mathbf{P}^t$ and the process is aperiodic, the stationary distribution π_r exists and is given by $\pi^* \cdot P_0 \cdot \ldots \cdot P_{r-1}$, since it satisfies the balance equation (1):

$$\begin{aligned} \pi_r &= \pi^* \cdot P_0 \cdot \ldots \cdot P_{r-1} \\ \pi_r &= \pi^* \cdot \mathbf{P} \cdot P_0 \cdot \ldots \cdot P_{r-1} \\ \pi_r &= \pi^* \cdot P_0 \cdot \ldots \cdot P_{r-1} \cdot P_r \cdot P_{r+1} \cdot \ldots \cdot P_{T-1} \cdot P_0 \cdot \ldots \cdot P_{r-1} \\ \pi_r &= \pi_r \cdot (P_r \cdot P_{r+1} \cdot \ldots \cdot P_{T-1} \cdot P_0 \cdot P_1 \cdot \ldots \cdot P_{r-1}). \end{aligned} \tag{1}$$

2.4. Persistence

The persistence of a given state s_i is related with the number of steps the system consecutively remains in this state. In the time-homogeneous case, it follows a geometric distribution with expected value $(1 - p_{i,i})^{-1}$ and is denoted by τ. Anastasiou and Tsekos [1] showed that it is possible to determine the expected time that a Markov chain stays consecutively inside a given subset of states using a simple closed-form expression. For example, in wind power applications, a typical subset of interest could contain all states corresponding to power production above a given threshold. To compute this estimate, the states are renumbered, *s*.th. they can be partitioned into two disjoint subsets: $\mathcal{A} = \{s_\nu, \ldots, s_n\}$ containing the states of interest; and $\overline{\mathcal{A}} = \{s_0, \ldots, s_{\nu-1}\}$, its complement. Then, the transition matrix is rearranged into the following block structure:

$$P = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{bmatrix} p_{0,0} & \cdots & p_{0,\nu-1} & p_{0,\nu} & \cdots & p_{0,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ p_{\nu-1,0} & \cdots & p_{\nu-1,\nu-1} & p_{\nu-1,\nu} & \cdots & p_{\nu-1,n} \\ p_{\nu,0} & \cdots & p_{\nu,\nu-1} & p_{\nu,\nu} & \cdots & p_{\nu,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ p_{n,0} & \cdots & p_{n,\nu-1} & p_{n,\nu} & \cdots & p_{n,n} \end{bmatrix},$$

where the first and last block of rows and columns correspond to the states in subset $\overline{\mathcal{A}}$ and $\mathcal{A}$, respectively. The expected value of persistence, i.e. the expected number of time steps the Markov process consecutively remains in the subset $\mathcal{A}$ once it is entered, is given by:

$$E\{\tau\} = \frac{\pi_{\mathcal{A}} 1_{\mathcal{A}}}{\pi_{\mathcal{A}} C 1_{\mathcal{B}}},$$

where $\pi_{\mathcal{A}}$ is the stationary probability distribution of the states in subset $\mathcal{A}$ and $1_{\mathcal{A}}$ is the unit column vector of size $(n - \nu + 1) \times 1$ [1].

For the time-inhomogeneous case, persistence τ_t is defined as the number of time steps the Markov process is expected to remain in a state (set of states), once it is entered at time t. For a cyclic Markov process, the persistence τ_t is equal for all t that are congruent modulo T. Thus, it is only necessary to compute the persistence for τ_r, $r = 0, \ldots, T-1$. This can be achieved by adapting the derivation of equation (2.4), provided by Anastasiou and Tsekos [1], to time-variant cyclic Markov processes.

After renaming, *s*.th. the subset of interest is $\mathcal{A} = \{s_\nu, \ldots, s_n\}$, the states of each transition matrix P_r are rearranged as in equation (2.4).

$$P = \begin{pmatrix} A_r & B_r \\ C_r & D_r \end{pmatrix} = \begin{bmatrix} p_{0,0}(r) & \cdots & p_{0,\nu-1}(r) & p_{0,\nu}(r) & \cdots & p_{0,n}(r) \\ \vdots & & \vdots & \vdots & & \vdots \\ p_{\nu-1,0}(r) & \cdots & p_{\nu-1,\nu-1}(r) & p_{\nu-1,\nu}(r) & \cdots & p_{\nu-1,n}(r) \\ p_{\nu,0}(r) & \cdots & p_{\nu,\nu-1}(r) & p_{\nu,\nu}(r) & \cdots & p_{\nu,n}(r) \\ \vdots & & \vdots & \vdots & & \vdots \\ p_{n,0}(r) & \cdots & p_{n,\nu-1}(r) & p_{n,\nu}(r) & \cdots & p_{n,n}(r) \end{bmatrix},$$

The probability of τ_r to be equal to l is given as:

$$\begin{aligned} Pr(\tau_r = l) &= Pr(X_r \in \mathcal{A}, \ldots, X_{r+l} \in \mathcal{A}, X_{r+l+1} \notin \mathcal{A} | X_r \in \mathcal{A}, X_{r-1} \notin \mathcal{A}) \\ &= \sum_{i \in \mathcal{A}} Pr(X_r = i | X_r \in \mathcal{A}, X_{r-1} \notin \mathcal{A}) \cdot Pr(X_s \in \mathcal{A}, r < s \\ &\leq l, X_{r+l+1} \notin \mathcal{A} | X_r = i) \\ &= \sum_{i \in \mathcal{A}} \sum_{k \notin \mathcal{A}} \sum_{j \in \mathcal{A}} \tilde{\pi}_r(i) \cdot \mathcal{P}_{i,j}(r, l-1, \mathcal{A}) \cdot p_{j,k}(r+l) \end{aligned} \quad (2)$$

with

$$\begin{aligned} \mathcal{P}_{i,j}(r, l, \mathcal{A}) &= Pr(X_{r+l} = j, X_k \in \mathcal{A}, 0 < k < l | X_r = i) \\ &= D_r \cdot \ldots \cdot D_{r+l-1} \cdot C_{r+l} \cdot 1_{\overline{\mathcal{A}}} \end{aligned}$$

and

$$\begin{aligned} \tilde{\pi}_r(i) &= Pr(X_r = i | X_r \in \mathcal{A}, X_{r-1} \notin \mathcal{A}) \\ &= \frac{\sum_{j \notin \mathcal{A}} Pr(X_r = i | X_{r-1} = j) \cdot Pr(X_{r-1} = j)}{\sum_{i \in \mathcal{A}} \sum_{j \notin \mathcal{A}} Pr(X_r = i | X_{r-1} = j) \cdot Pr(X_{r-1} = j)} \\ &= \frac{\sum_{j \notin \mathcal{A}} \pi_{r-1}(j) p_{j,i}(r-1)}{\sum_{k \in \mathcal{A}} \sum_{j \notin \mathcal{A}} \pi_{r-1}(j) p_{j,k}(r-1)}, \quad i \in \mathcal{A}, \end{aligned} \quad (3)$$

where $\pi_r(j)$ is the long-term probability of occurrence (stationary probability) of state j at time of the day r; also note that $\pi_t(j) = \pi_r(j)$ for $t = mr$, $\forall m \in \mathbb{N}$. Equation (3) can be rewritten in the matrix form to include all states in the subset $\mathcal{A}$:

$$\tilde{\pi}_r(\mathcal{A}) = \frac{\pi_{r-1}\left(\overline{\mathcal{A}}\right) \cdot B_{r-1}}{\pi_{r-1}\left(\overline{\mathcal{A}}\right) \cdot B_{r-1} \cdot 1_{\mathcal{A}}},$$

where $1_{\mathcal{A}}$ is a unit vector of dimension $(n - \nu + 1) \times 1$ and $\pi_{r-1}(\overline{\mathcal{A}})$ is a vector of dimensions $1 \times \nu$, whose elements are the stationary probabilities of the states in the set $\overline{\mathcal{A}}$ at time of the day $r - 1$.

Thus, equation (2) can be rewritten as:

$$\begin{aligned} Pr(\tau_r = l) &= \tilde{\pi}_r(\mathcal{A}) \cdot D_r \cdot \ldots \cdot D_{r+l-1} \cdot C_{r+l} \cdot 1_{\overline{\mathcal{A}}} \\ &= \frac{\pi_{r-1}\left(\overline{\mathcal{A}}\right) \cdot B_{r-1}}{\pi_{r-1}\left(\overline{\mathcal{A}}\right) \cdot B_{r-1} \cdot 1_{\mathcal{A}}} \cdot D_r \cdot \ldots \cdot D_{r+l-1} \cdot C_{r+l} \cdot 1_{\overline{\mathcal{A}}}. \end{aligned}$$

The expected value of persistence at time r can then be derived as:

$$E(\tau_r) = \sum_{l=1}^{\infty} l \cdot \frac{\pi_{r-1}\left(\overline{\mathcal{A}}\right) \cdot B_{r-1}}{\pi_{r-1}\left(\overline{\mathcal{A}}\right) \cdot B_{r-1} \cdot 1_{\mathcal{A}}} \cdot D_r \cdot \ldots \cdot D_{r+l-1} \cdot C_{r+l} \cdot 1_{\overline{\mathcal{A}}}$$

Making use of the cyclicity of the Markov process, this can be expressed as:

$$E(\tau_r) = \sum_{l=1}^{\infty} l \cdot \frac{\pi_{r-1}\left(\overline{\mathcal{A}}\right) \cdot B_{r-1}}{\pi_{r-1}\left(\overline{\mathcal{A}}\right) \cdot B_{r-1} \cdot 1_{\mathcal{A}}} \cdot \mathcal{D}^m \cdot D_r \cdot \ldots \cdot D_{r+r_l-1} \cdot C_{r+r_l} \cdot 1_{\overline{\mathcal{A}}}$$

where $\mathcal{D} = D_r \cdot \ldots \cdot D_T \cdot D_{T+1} \cdot \ldots \cdot D_{r-1}$ and $l = mT + r_l$.

It can be seen that the sum converges after splitting it into T partial sums, one for each time of the day r. For each partial sum, the only term not constant is the matrix power $\mathcal{D}^m$, which converges because all eigenvalues of $\mathcal{D}$ are smaller than 1. The infinite sum for the expected value of persistence at time r can be approximated to an arbitrary degree of accuracy ε by defining

$$f_l = l \cdot \frac{\pi_{r-1}\left(\overline{\mathcal{A}}\right) \cdot B_{r-1}}{\pi_{r-1}\left(\overline{\mathcal{A}}\right) \cdot B_{r-1} \cdot 1_{\mathcal{A}}} \cdot \mathcal{D}^m \cdot D_r \cdot \ldots \cdot D_{r+r_l-1} \cdot C_{r+r_l} \cdot 1_{\overline{\mathcal{A}}}.$$

and successively adding f_l, $l = 0, 1, \ldots, L$ until the difference between two consecutive sums is smaller than ε, i.e. until $|f_L| < \varepsilon$.

3. Parameter estimation

3.1. Time-invariant Markov chain

The common approach to estimate the Markov chain transition matrix P is through the optimization of a constrained maximum likelihood function, which describes the realization probability of a given dataset [2]. For a sequence of M states $\{X_0 = s_{i_0}, \ldots, X_M = s_{i_M}\}$ with $s_{i_0}, \ldots, s_{i_M} \in S$ and $i_0, \ldots, i_M \in \{0, \ldots, n\}$, its probability can be computed as $Pr\{X_0 = s_{i_0}\} p_{i_0,i_1} p_{i_1,i_2} \cdot \ldots \cdot p_{i_{M-1},i_M}$. Since the term $Pr\{X_0 = s_{i_0}\}$ is constant, given a set of observed state transitions $\mathcal{S}$, it is possible to estimate P by maximizing the likelihood

$$\mathrm{OF}_1 = \prod_{(i,j) \in \mathcal{S}} p_{i,j}, \quad (4)$$

where a transition is described by an ordered pair (i, j) indicating the origin and the destination of the transition. In practice, instead of maximizing OF_1 with respect to the $p_{i,j}$ variables it is preferable to minimize the negative log-likelihood function, i.e. $-\log(\mathrm{OF}_1)$, since it transforms the original mathematical programming problem into an equivalent one that is convex and, thus, has a unique solution [4]. The overall optimization problem is formulated as follows:

$$\begin{aligned} \min \quad & -\sum_{(i,j) \in \mathcal{S}} \log\left(p_{i,j}\right) \\ \text{subject to} \quad & p_{i,j} \geq 0 \quad \forall i, j = 0, \ldots, n \\ & \sum_j p_{i,j} = 1 \quad \forall i = 0, \ldots, n \end{aligned}$$

The constraints ensure non-negativity of the transition probabilities and that they sum up to 1 for each row of the transition matrix.

3.2. Cyclic time-variant Markov process

The goal of this time-variant Markov process is to get a model that accurately reproduces the long-term behavior while considering the daily patterns observed in the data. Thus, the proposed objective function combines two maximum likelihood estimators: the first term maximizes the likelihood of the cycle-average probability; and, the second term maximizes the likelihood of the time-dependent probability. The final optimization problem is transformed into a convex one using the negative logarithm of the objective function. This section provides a detailed description of the objective function, the parameterization of the time-variant probability functions, and the constraints that must be added to the optimization problem to ensure its Markov properties.

3.2.1. Objective function

The transition probabilities are considered to be time-variant and cyclic with a period of T, i.e. for each time of the day r (=0,

…, $T-1$) there is a different transition matrix P_r. In this paper, the time-dependent transition probabilities $p_{i,j}(r)$ are modeled by Bernstein polynomials. This has several advantages: a) a polynomial representation of the transition probabilities leads to a convex objective function and constraints, i.e. the optimization problem has a unique solution; b) a polynomial representation allows to decrease the number of variables in the optimization problem: for each transition, instead of T variables only $k+1$ are needed for a k order polynomial; c) Bernstein polynomials are non-negative, which simplifies probability modeling, when compared to other polynomial bases; and d) they have the convex hull property, which, combined with de Casteljau algorithm, allows to easily write probability boundary conditions.

Bernstein polynomials are linear combinations of Bernstein basis polynomials $b_{\mu,k}(z)$, $z\in[0,1]$. The $k+1$ Bernstein basis polynomials of order k are defined as:

$$b_{\mu,k}(z) = \binom{k}{\mu} z^{\mu}(1-z)^{k-\mu}$$

with $\mu = 0, \ldots, k$ and $(k\mu)$ the binomial coefficient. Thus, the transition probabilities $p_{i,j}(z)$ are described by

$$p_{i,j}(z) = \sum_{\mu=0}^{k} \beta_{\mu}^{i,j} b_{\mu,k}(z),$$

with $\beta_{\mu}^{i,j}\in\mathbb{R}$ and $z = r/T$, since the polynomial variable has to be scaled, s.th. it is between 0 and 1.

To maximize the likelihood of the time-dependent transition probabilities given the data, the objective function must consider the time of the day z when the transition happens. Therefore, the objective function introduced in Eq. (4) becomes $\sum_{(i,j)_z\in\mathcal{S}_z} \log(p_{i,j}(z))$, where $\mathcal{S}_z$ is the set of observed transitions together with the time z when they happen. This objective function allows to compute the intra-cycle transition probability functions, and thus to represent the daily patterns present in the data.

A second term is added to this function, namely $\sum_{(i,j)\in\mathcal{S}} \log(p_{i,j}^{\text{avg}})$, where $\mathcal{S}$ is the set of transitions observed in the data as defined in Section 3.1 and $p_{i,j}^{\text{avg}}$ is the cycle-average (daily) probability of transition from state s_i to s_j. It can be computed as follows:

$$p_{i,j}^{\text{avg}} = \frac{1}{1-0}\int_0^1 p_{i,j}(z)\mathrm{d}z = \int_0^1 \sum_{\mu=0}^{k} \beta_{\mu}^{i,j} b_{\mu,k}(z)\mathrm{d}z$$
$$= \sum_{\mu=0}^{k} \beta_{\mu}^{i,j} \int_0^1 b_{\mu,k}(z)\mathrm{d}z = \frac{1}{k+1}\sum_{\mu=0}^{k} \beta_{\mu}^{i,j}$$

This second term is the maximum likelihood estimator for the daily average probability and its addition to the objective function increases the consistency of the long-term behavior of the Markov process with the data. Therefore, the overall objective function OF_2 is given by:

$$\text{OF}_2 = -\sum_{(i,j)\in\mathcal{S}} \log\left(p_{i,j}^{\text{avg}}\right) - \sum_{(i,j)_z\in\mathcal{S}_z} \log\left(p_{i,j}(z)\right)$$
$$= -\sum_{(i,j)\in\mathcal{S}} \log\left(\frac{1}{k+1}\sum_{\mu=0}^{k}\beta_{\mu}^{i,j}\right) - \sum_{(i,j)_z\in\mathcal{S}_z} \log\left(\sum_{\mu=0}^{k}\beta_{\mu}^{i,j} b_{\mu,k}(z)\right)$$

and minimization is performed with respect to the coefficients $\beta_{\mu}^{i,j}$ (model parameters).

3.2.2. Constraints

The estimation of the model parameters requires the transition probability functions to comply with several constraints, to ensure:

- $\mathcal{C}^0$- and $\mathcal{C}^1$-continuity at $z = 0$,
- row-stochasticity of the transition matrices at every time of the day z and
- that the transition probability functions are non-negative and bounded by 1.

Thus, to complete the specification of the optimization problem this section explains all the necessary constraints required for the model parameters to describe a cyclic Markov process.

3.2.2.1. Periodicity. The transition probability functions are modeled using Bernstein polynomials, which are smooth, i.e. $\mathcal{C}^{\infty}$-continuous functions. In general, the values at both ends of their domain (0 and 1) need not be equal. Thus, to avoid sudden changes in the value and slope of each probability function between the cycles, two constraints are added to ensure $\mathcal{C}^0-$ and $\mathcal{C}^1$-continuity. Another reason is the arbitrariness of the cycle starting position, which affects the position of the discontinuity if these conditions are not used.

The first constraint is $p_{i,j}(0) = p_{i,j}(1)$. Since $b_{\mu,k}(0) = \delta_{\mu,0}$ and $b_{\mu,k}(1) = \delta_{\mu,k}$ the constraint can be reformulated as $\beta_0^{i,j} = \beta_k^{i,j}$, where δ is the Kronecker delta. The second constraint is added to ensure $\mathcal{C}^1$-continuity, i.e. $\mathrm{d}p_{i,j}/\mathrm{d}z(0) = \mathrm{d}p_{i,j}/\mathrm{d}z(1)$. The first derivative of a Bernstein basis polynomial can be written as a combination of two polynomials of lower degree:

$$\frac{\mathrm{d}b_{\mu,k}}{\mathrm{d}z}(z) = k\left(b_{\mu-1,k-1}(z) - b_{\mu,k-1}(z)\right)$$

Thus, the first derivative of a transition probability $p_{i,j}(z)$ is given by:

$$\frac{\mathrm{d}p_{i,j}}{\mathrm{d}z}(z) = k\left(\sum_{\mu=1}^{k}\left(\beta_{\mu}^{i,j} - \beta_{\mu-1}^{i,j}\right) b_{\mu-1,k-1}(z) - \beta_k^{i,j} b_{k,k-1}(z)\right)$$

Hence, using $b_{\mu,k}(0) = \delta_{\mu,0}$ and $b_{\mu,k}(1) = \delta_{\mu,k}$ as well as the first constraint $\beta_0^{i,j} = \beta_k^{i,j}\ \forall i,j = 0,\ldots,n$, the constraint $\mathrm{d}p_{i,j}/\mathrm{d}z(0) = \mathrm{d}p_{i,j}/\mathrm{d}z(1)$ reduces to the following linear constraint:

$$\beta_k^{i,j} = \beta_0^{i,j} = 0.5\left(\beta_1^{i,j} + \beta_{k-1}^{i,j}\right)$$

3.2.2.2. Row stochasticity of transition matrices. To ensure row stochasticity of the time-variant transition matrices, it is necessary to ensure that $\sum_j p_{i,j}(z) = 1$ for all i and z. Since the Bernstein basis polynomials of order k form a partition of unity, i.e.

$$\sum_{\mu=0}^{k} b_{\mu,k}(z) = 1$$

the constraint can be re-written as a linear combination of the polynomial coefficients:

$$\sum_j p_{i,j}(z) = 1 \Leftrightarrow \sum_j \sum_{\mu=0}^{k} \beta_{\mu}^{i,j} b_{\mu,k}(z)$$
$$= 1 \Leftrightarrow \sum_j \beta_{\mu}^{i,j} = 1$$

3.2.2.3. Non-negative transition probabilities are bounded by 1. The most straightforward way to implement this constraint is to

add two inequalities for each time of the day and each transition probability $p_{i,j}$, i.e.

$$\begin{aligned} p_{i,j}(z) &\geq 0 \quad \forall i,j,z \\ p_{i,j}(z) &\leq 1 \quad \forall i,j,z \end{aligned} \tag{5}$$

However, this constraint significantly increases the problem size, since it requires $2 \cdot T \cdot n^2$ inequalities. An alternative constraint can be formulated by using the convex hull property of the Bernstein polynomials. This constraint makes the overall optimization problem size smaller, but is more restrictive.

Every Bernstein polynomial $\sum_{\mu=0}^{k} \beta_\mu b_{\mu,k}(z)$ always lies in the convex hull defined by its control points $(k/\mu, \beta_\mu)$, $\mu = 0, \ldots, k$. Thus the constraint

$$0 \leq p_{i,j}(z) \leq 1 \quad \forall i,j = 0,\ldots,n$$

can be reformulated in terms of the polynomial coefficients as

$$0 \leq \beta_\mu^{i,j} \leq 1 \quad \forall \mu = 0,\ldots,k \tag{6}$$

Since constraint (6) is a sufficient but not necessary condition for constraint (5), the reformulation leads to a more restrictive overall minimization problem, i.e. the optimum objective function value is always higher or equal when compared with the problem with original constraint (5). The convex hull bound of Bernstein polynomials can be tightened by subdivision, i.e. by subdividing the domain in two regions and finding new control points $\beta_0^{i,j}(1), \ldots, \beta_k^{i,j}(1)$ and $\beta_{k+1}^{i,j}(1), \ldots, \beta_{2k}^{i,j}(1)$ such that the function output remains unchanged. With each subdivision, the control points form a tighter bound around the polynomial and thus the polynomial coefficients can assume values in a wider range. The new control points represent the polynomial restricted to the two sub-intervals $[0, z^*]$ and $[z^*, 1]$, where $z^* \in [0, 1]$ is the cutting point of the division. For simplicity, z^* is fixed to 0.5 in all transition probabilities. The new control points can be determined by linear combinations of the original control points $\beta_0^{i,j}, \ldots, \beta_k^{i,j}$. This can be performed efficiently using de Casteljau algorithm, which in matrix form is given as:

$$\begin{aligned} \begin{pmatrix} \beta_0^{i,j}(1) \\ \vdots \\ \beta_k^{i,j}(1) \end{pmatrix} &= \begin{pmatrix} b_{0,0}(z^*) & 0 & \cdots & 0 \\ b_{0,1}(z^*) & b_{1,1}(z^*) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b_{0,k}(z^*) & b_{1,k}(z^*) & \cdots & b_{k,k}(z^*) \end{pmatrix} \begin{pmatrix} \beta_0^{i,j} \\ \vdots \\ \beta_k^{i,j} \end{pmatrix} \\ &= C_l \begin{pmatrix} \beta_0^{i,j} \\ \vdots \\ \beta_k^{i,j} \end{pmatrix} = C_l \cdot \beta^{ij} \end{aligned} \tag{7}$$

and

$$\begin{aligned} \begin{pmatrix} \beta_{k+1}^{i,j}(1) \\ \vdots \\ \beta_{2k}^{i,j}(1) \end{pmatrix} &= \begin{pmatrix} b_{0,k}(z^*) & b_{1,k}(z^*) & \cdots & b_{k,k}(z^*) \\ 0 & b_{0,k-1}(z^*) & \cdots & b_{k-1,k-1}(z^*) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & b_{0,0}(z^*) \end{pmatrix} \begin{pmatrix} \beta_0^{i,j} \\ \vdots \\ \beta_k^{i,j} \end{pmatrix} \\ &= C_r \begin{pmatrix} \beta_0^{i,j} \\ \vdots \\ \beta_k^{i,j} \end{pmatrix} = C_r \cdot \beta^{i,j} \end{aligned} \tag{8}$$

The subdivision can be applied recursively to further improve the convex bound around the polynomial. The corresponding coefficients are computed by applying equations (7) and (8) to the left and right coefficient vectors. Defining $C = (C_l, C_r)^T$ and I_z as the identity matrix of dimension $z \times z$, the coefficients $\beta^{i,j}(w) = \left(\beta_0^{i,j}(w), \ldots, \beta_{2^w k}^{i,j}(w)\right)$ after w subdivisions can be obtained by:

$$\beta^{i,j}(w) = (C \otimes I_{2^{w-1}}) \cdot (C \otimes I_{2^{w-2}}) \cdot \ldots \cdot (C \otimes I_2) \cdot (C \otimes I_1) \cdot \beta^{i,j}$$

where $\otimes$ denotes the Kronecker product. The number of inequalities needed for the implementation of this constraint is $(k+1) \cdot 2^{\omega+1} \cdot n^2$. Thus, its use only makes sense if it decreases the problem size, i.e. for a number of subdivisions ω such that $(k+1) \cdot 2^{\omega+1} \leq T$.

3.2.3. Problem formulation

The overall optimization problem to be solved for the estimation of the transition probability coefficients $\beta_\mu^{i,j}$ can be written as:

$$\min_{\beta_\mu^{i,j}} -\sum_{(i,j)\in\mathcal{S}} \log\left(\frac{1}{k+1}\sum_{\mu=0}^{k}\beta_\mu^{i,j}\right) - \sum_{(i,j)_z\in\mathcal{S}_z} \log\left(\sum_{\mu=0}^{k}\beta_\mu^{i,j} b_{\mu,k}(z)\right) \tag{9}$$

$$\text{subject to } \beta_0^{i,j} = \beta_k^{i,j} \quad \forall i,j = 0,\ldots,n \tag{10}$$

$$\beta_0^{i,j} = 0.5\left(\beta_1^{i,j} + \beta_{k-1}^{i,j}\right) \quad \forall i,j = 0,\ldots,n \tag{11}$$

$$\sum_j \beta_\mu^{i,j} = 1 \quad \forall i = 0,\ldots,n; \forall \mu = 0,\ldots,k \tag{12}$$

$$\beta^{i,j}(w) \leq 1 \quad \forall i,j = 0,\ldots,n \tag{13}$$

$$0 \leq \beta^{i,j}(w) \quad \forall i,j = 0,\ldots,n \tag{14}$$

where w is the number of subdivisions and k is the order of the Bernstein polynomials, which have to be specified. The objective function [9] is a combination of two negative log-likelihood functions to ensure the Markov process captures both the daily patterns and the long-term behavior of the original data. The optimization is performed with respect to several constraints: constraints (10) and (11) ensure $\mathcal{C}^0$- and $\mathcal{C}^1$-continuity at $z = 0$. The row-stochasticity of the transition matrix is ensured by constraint (12). The last two constraints (13) and (14) bound the transition probabilities between 0 and 1.

It is expected that the objective function decreases with the polynomial order and the number of subdivisions. The parameters of the Markov chain model $\beta_\mu^{i,j}$ are estimated by solving the optimization problem using a rigorous numerical solver. The model was formulated making use of the casadi computation framework [3] and the optimization was performed by IPOPT (Interior Point OPTimizer), a nonlinear interior-point solver [27], which ensures convergence to the global optimum in the case of convex optimization problems.

4. Application of the cyclic Markov process to wind turbine modeling

4.1. The data

The data for this study was obtained from a wind power turbine in a wind park located in a mountainous region in Portugal. The time series consists of a three-year period (2009–2011) of historical data gotten from the turbine data logger. The sampling time of 10 min leads to 144 samples each day. The dataset comprises three variables, wind power, speed and direction (nacelle orientation). The wind speed information was collected from the anemometer placed in the wind turbine hub. Due to confidentiality, wind power and speed data values are reported as a fraction of the rated power and the cut-out speed, respectively.

4.2. Markov process state definition

Discrete Markov models require the definition of the states when applied to describe continuous variables. This work proposes to characterize the wind turbine states using three different variables: wind power, speed and direction. As such, each state is defined by all the points inside a polyhedron in three-dimensional space.

Fig. 1 presents all data observations and the state partitions projected into: a) the wind direction and speed plane; and, b) the wind power and speed plane. As expected, the observations projected into the wind power and speed plane define the characteristic power curve of the wind turbine. It shows the three operational regions of a wind turbine: a) below the cut-in speed no power is produced; b) between cut-in and rated wind speed the power increases proportionally to the cube of wind speed; c) at wind speeds between the rated and the cut-out wind speed, the turbine control system limits the power output to a constant value. In the wind direction and speed plane, data is widely scattered and shows the dominant wind patterns at the site. Three accumulation regions can be identified: one for low wind speeds, centered on 0.25, which is the mode of the wind speed, and two defining the dominant wind directions around 100° and 300°.

The data space is discretized unevenly to get a good resolution of the high-slope region of the power curve. In a previous work [13], this partition was used in a time-homogeneous Markov chain and proved to lead to an accurate representation of the original data. The wind direction and power are divided by an equally spaced grid leading to 12 ($\{d_1, \ldots, d_{12}\}$) and 20 ($\{p_1, \ldots, p_{20}\}$) classes, respectively. The wind speed is divided as follows: values below the cut-in speed define one class sp_1; between the cut-in and rated wind speed the discretization is narrowed by selecting 10 classes ($\{\mathrm{sp}_2, \ldots, \mathrm{sp}_{11}\}$); and between the rated and cut-out wind speed discretization is widened and 4 classes ($\{\mathrm{sp}_{12}, \ldots, \mathrm{sp}_{15}\}$) are defined. Datapoints with wind speed above the cut-out wind speed are discarded. The complete state set is constructed by listing all possible combinations of the classes of each variable. Due to physical constraints between the variables, most of the states are empty (Fig. 1(left)) and can be discarded. This reduces the number of states from 3840 to 778, for this turbine.

4.3. Additional transitions to promote a single communication class

The solution of the optimization problem described in Section 3 comprises a set of transition matrices P_r, $r \in \{0, \ldots, 143\}$. However, the constraints in the optimization problem definition do not force the matrix $\mathbf{P} = P_0 \cdot \ldots \cdot P_{143}$ to be irreducible and thus the Markov process to have a single communication class. So, during data synthesis, the Markov process can get "trapped" within a communication class. To induce a single communication class in the Markov process, additional transition counts are introduced into the data. The goal is to add a small set of transitions to promote state connectivity without distorting the original data. Thus, the set is composed of transitions that connect neighboring states in the state space, since those are the ones most likely to occur.

For a state $s_i = (p_l, \mathrm{sp}_p, d_q)$, its neighborhood V is defined as

$$V(s_i) := \left\{ \left(p_{\mathrm{ll}}, \mathrm{sp}_{\mathrm{pp}}, d_{\mathrm{qq}} \right) : \mathrm{ll} \in \{l-1, l, l+1\}, \mathrm{pp} \in \times \{p-1, p, p+1\}, \mathrm{qq} \in \{q-1, q, q+1\} \right\} \setminus s_i. \tag{15}$$

It should be noted that, unlike power and speed, direction is a circular variable, e.g. states d_0 and d_{11} are considered neighbors. If a neighbor state $s_j \in V(s_i)$ is present in the dataset, a transition (i, j) is added to the set of extra transitions $\mathcal{S}_{\mathrm{E}}$. For this dataset, originally consisting of 150,601 transitions, 13,610 transitions are added.

The extra transitions must be considered to happen at an unknown time of the day z. Thus, they can only be accounted for in the objective function term without time information, i.e. only in the time-variant part of the objective function. This directly affects the values for $p_{i,j}^{\mathrm{avg}}$ and, indirectly, the model parameters. Since the aim is to cause a minimal impact on the transition probabilities, the additional term is weighed by a factor $\omega < 1$ to directly control its influence. In this work it is fixed to 0.05. Thus, the following term is added to the objective function:

$$-\omega \cdot \sum_{(i,j) \in \mathcal{S}_E} \log\left(p_{i,j}^{\mathrm{avg}}\right) = -\omega \cdot \sum_{(i,j) \in \mathcal{S}_E} \log\left(\frac{1}{k+1} \sum_{\mu=0}^{k} \beta_{\mu}^{i,j} \right) \tag{16}$$

Although the use of the extra transition set does not ensure the time-variant Markov process to have a single communication class, results show a decrease of the number of communication classes from 13 to 1 in this dataset.

5. Simulation of wind power, speed and direction time series

To simulate wind power, speed and direction time series the method described by Sahin and Sen [22] is adapted to the cyclic time-variant Markov model as follows. First, the cumulative probability transition matrices P_r^{cum} with $P_r^{\mathrm{cum}}(i,j) = \sum_{k=0}^{j} p_{i,k}(r)$ are computed. Then an initial state s_i, i.e. $X_0 = s_i$, is randomly selected. A new datapoint X_{t+1} is generated by uniformly selecting a random number ε between zero and one. The corresponding state s_{new} ($X_{t+1} = s_{\mathrm{new}}$) is chosen such that the probability of reaching it from the current state s_i is bigger than ε, i.e. such that $P_{r_t}^{\mathrm{cum}}(i, \mathrm{new}) \geq \varepsilon$.

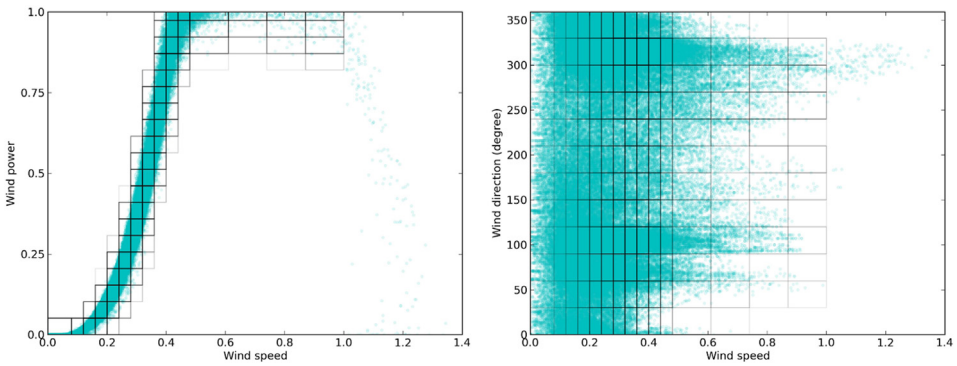


Fig. 1. Representation of all datapoints projected into the: a) wind direction and speed plane (left); and, b) wind power and speed plane (right). Each rectangle is the projection of a state polyhedron into the two planes. Overall, they define the final state partition for the three-dimensional variable space.

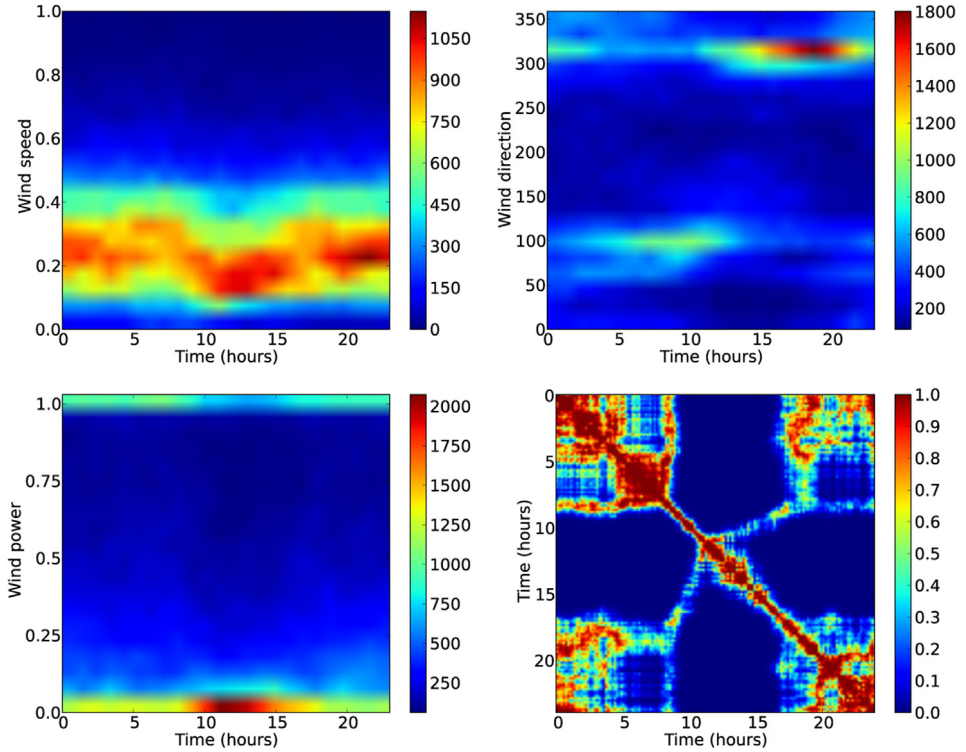


Fig. 2. Two-dimensional histograms of the original time series data: speed–time (top left), direction–time (top right) and power–time (bottom left). The subfigure on the bottom right shows the p-value of the Kolmogorov–Smirnov test used to compare the wind speed distribution on the different times of the day.

Based on this discrete state sequence, a real value for the wind power/speed/direction variables is generated by sampling each state partition uniformly.

6. Results and discussion

6.1. Daily patterns in the data

The wind power, speed and direction time series clearly show a daily time-dependent behavior.

Fig. 2 shows that, on average, the turbine does not produce power between 10 am and 3 pm. In this time interval, low wind speeds (0.1–0.25) are the most likely events. There are two dominant wind directions: around 100° and 300°. Moreover, they occur at specific times of the day; between 5 and 10 am, the wind typically blows from the 100° direction, the rest of the day from 300°.

To assess whether these two dominant directions might be due to summer/winter seasonality, the dataset was divided in two subsets, one covering the period from April to September and the other from October to March. The histogram analysis shows that both, summer and winter subset, have the same two dominant directions (figures not shown). Thus, it was concluded that the time-dependent pattern is not induced by this seasonality.

Fig. 2 bottom-right plot shows the p-values of the Kolmogorov–Smirnov test applied to the wind speed distributions at different times of the day. The Kolmogorov–Smirnov test is a nonparametric test for the equality of continuous one-dimensional probability distributions. Thus, the high p-values around the diagonal illustrates that wind speed distributions for consecutive times of the day are similar. The same holds for wind speeds in the morning and evening, i.e. before 9 am and after 4:30 pm. The wind speed distribution between 10 am and 3 pm is clearly different.

6.2. Choice of polynomial order and number of subdivisions

The model introduced in Section 3.2 has two parameters that need to be defined: k, the order of the Bernstein polynomials used to model the transition probabilities; and w, the number of subdivisions used to tighten the convex hull that bounds the polynomials. To choose proper values for these parameters, different models were computed by varying $k = 4\ldots10$ and $w = 0\ldots3$. For each model, synthetic data was generated following the procedure described in Section 5 and compared with the real dataset.

Fig. 3 shows bar plots comparing the different models using four criteria: the objective function value, the daily average Jensen-Shannon distance between original and synthesized wind direction data, the number of inequalities in the problem formulation and the CPU (central processing unit) time spent in IPOPT solving the optimization problem. The Jensen–Shannon distance is the square root of the Jensen–Shannon divergence d_{js}, which, for two discrete probability distributions q_1 and q_2 is defined as:

$$d_{\mathrm{js}} = \frac{1}{2}\sum_i \frac{q_1(i)}{q_2(i)} \cdot q_1(i) + \frac{1}{2}\sum_i \frac{q_2(i)}{q_1(i)} \cdot q_1(i).$$

Comparing the models using the objective function value (Fig. 3 top left) shows a decrease of the objective function as the model order and the number of subdivisions increase. It can be seen that the impact of the number of subdivisions is higher for models with higher polynomial order. Moreover, the first subdivision has the highest impact since it leads to the highest decrease of the objective function value. The daily average of the Jensen–Shannon distance (Fig. 3 top right) decreases with the polynomial order until sixth order. The same behavior can be observed for the number of subdivisions: until the sixth order, the Jensen–Shannon distance

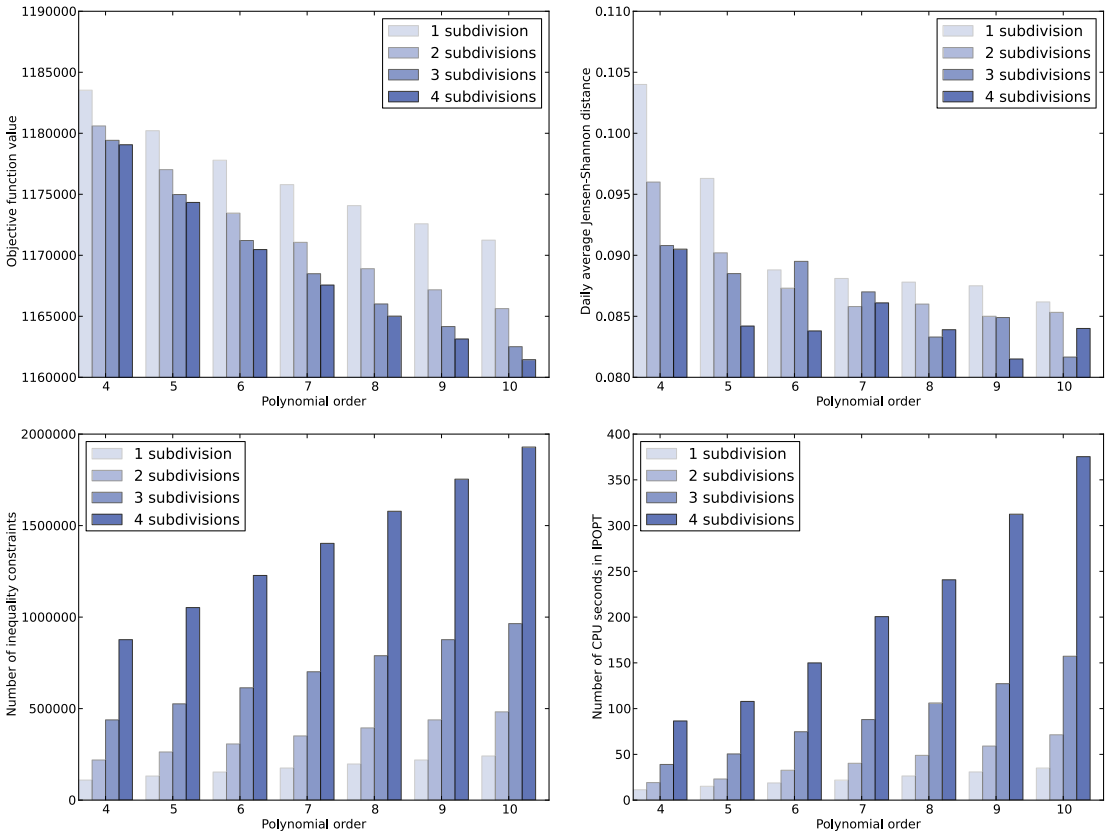


Fig. 3. Bar plots comparing the objective function value (top left), the daily average Jensen–Shannon distance between original and synthesized data (top right), the number of inequalities in the problem formulation (bottom left) and the CPU time spent in IPOPT solving the optimization problem (bottom right) of the tested models.

decreases with the number of subdivisions. The number of inequality constraints in the optimization problem as well as the number of CPU seconds spent in the solver show the expected behavior (Fig. 3 bottom). They increase linearly with the polynomial order and exponentially with the number of subdivisions. Based on these observations, a basis order of 6 with 2 subdivisions was chosen as the best trade-off between an accurate representation of the average daily patterns and computational costs.

6.3. *Capturing long-term statistics*

This section compares the main statistical properties derived from the original data with the ones derived from the data generated by the time-variant Markov model.

Fig. 4 compares the wind power (left), speed (middle) and direction (right) distribution of the original with the synthetic data generated using the Markov model. In general, the distributions are in close agreement. The wind power distribution is bimodal, with the modes located at the minimum and maximum power. It shows that the intermediary power levels are rather rare, for instance, the states corresponding to a power production between 0.4 and 0.9 have a low probability. The wind speed distribution follows the expected behavior, a single mode distribution with a long tail for the high wind speeds (Weibull distribution). The wind direction distribution is bimodal with the two modes at 100 and 300°, which are the prevailing wind directions at the turbine site (Fig. 2).

Fig. 5 shows two plots: on the left, the empirical 2D distribution of the wind power and direction computed from the data and, on the right, the same distribution computed using the data generated by the Markov model. Its comparison shows that the model captures the joint statistics for the wind power and direction from the data. It is possible to see the two dominant directions associated with high wind power production, namely the sectors from 100 to 120 and from 290 to 320°. Figs. 2 and 5 clearly

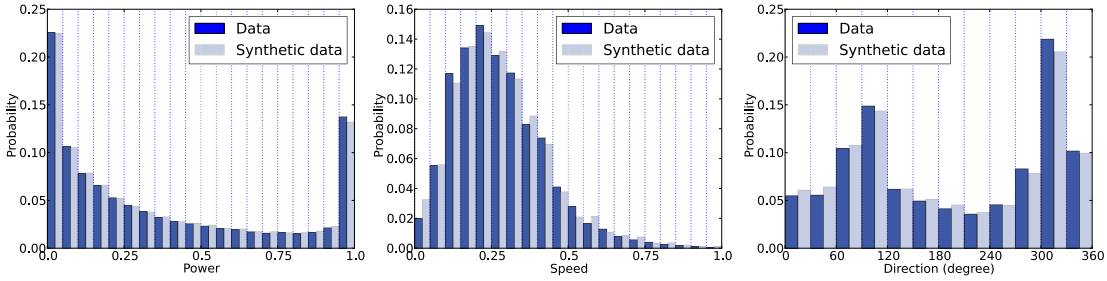


Fig. 4. Comparison of the probability distribution of wind power (left), wind speed (middle) and wind direction (right) of the original with the synthesized data.

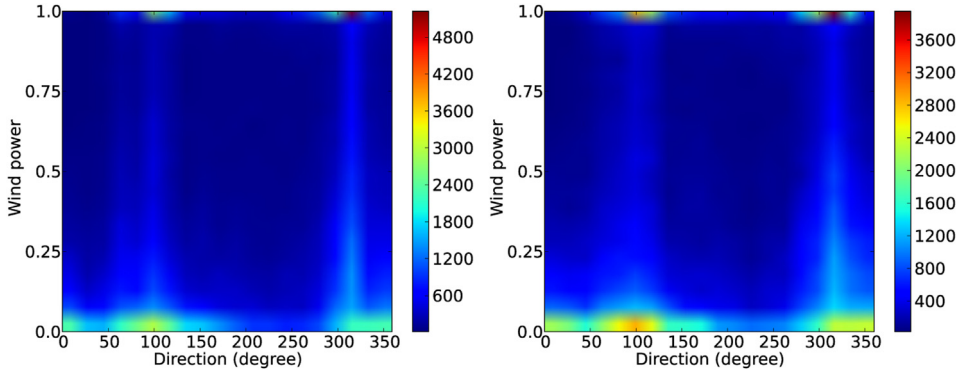


Fig. 5. Two dimensional power–time histograms of the original (left) and synthetic (right) time-series data.

demonstrate the capability of this Markov model to capture the combined characteristics of the wind power, speed and direction. The long-term behavior of the model is close to what is observed in the dataset.

6.4. Capturing time-dependent behavior

As shown in Section 6.1, the original data clearly exhibits a time-dependent behavior. To test, if the time-dependent Markov model

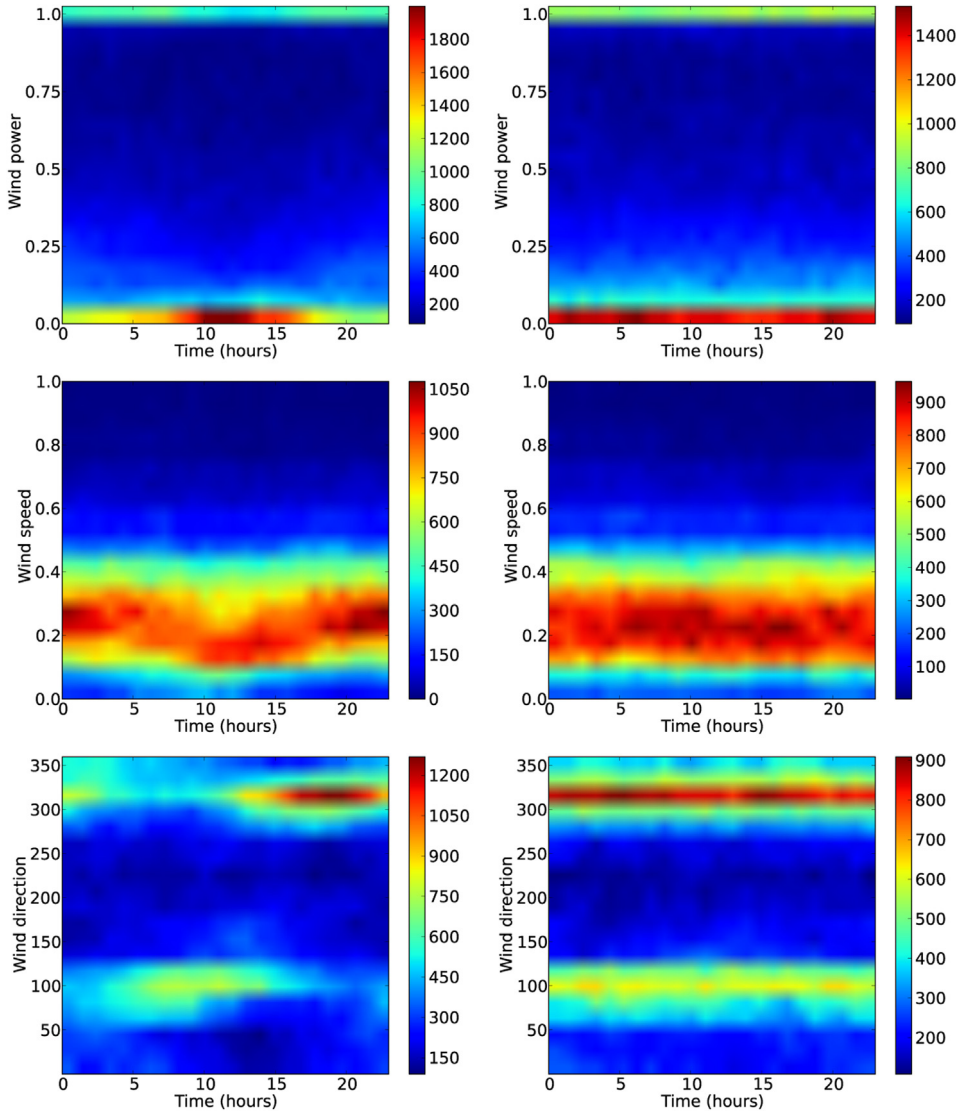


Fig. 6. Two dimensional histograms of the synthetic time-series data, generated with the time-variant Markov model (left) and the time-invariant Markov chain (right): power–time (top), speed–time (middle) and direction–time (bottom).

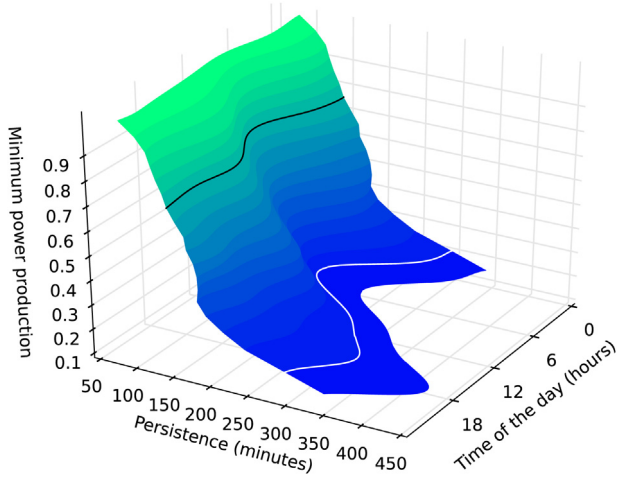


Fig. 7. Time-dependent persistence of power production above $i \cdot 0.05 \cdot p_{max}$, for $i = 1, \ldots, 19$. The lines highlight the time-dependent persistence, for two conditions: a) in white, for power production above $0.15 \cdot p_{max}$ (PUPP); and, b) in black, the power production above $0.7 \cdot p_{max}$ (PHPP).

can capture it, synthetic data was generated and the histograms compared to the ones of the original data. Moreover, to obtain a comparison with the "regular" way of data synthesis with Markov models, data was also generated from the time-invariant Markov chain.

The comparison of Figs. 2 and 6(first column) shows, that the time-variant Markov model is capable of reproducing the time-variant behavior of the data. Fig. 6(second column) presents the results of using a time-invariant Markov chain model, i.e. by using constant transition probability functions. As expected, each variable statistic distribution remains constant during the daily cycle.

6.5. Time-dependent persistence of production

The time-dependent Markov model allows to compute the persistence of power-production depending on the time of the day. Fig. 7 shows the time-dependent persistence of power production for different power levels ($i \cdot 0.05 \cdot p_{max}$, for $i = 1, \ldots, 19$). The persistence analysis is presented for two power production levels: a) PUPP (persistence of useful power production) defined as above $0.15 \cdot p_{max}$, i.e. the power level corresponding to the wind speed mode at the turbine site; and, PHPP (persistence of high power production), i.e. above $>0.7\ p_{max}$. It can be seen, that the higher the power level, the lower the persistence. Moreover, for all power levels, persistence is minimal between 5 and 10 am. PHPP is fairly constant throughout the day (dark line), the maximal differences are between 10 and 30 min, whereas PUPP reaches a maximum at around 5 pm (white line).

Since the data shows two different dominant directions (Fig. 4), Fig. 8 presents the persistence of power production conditioned to each dominant direction, i.e. for the direction sectors from 90° to 180° and 270° to 360°.

As expected, the persistence in both direction sectors is lower than the unconditioned persistence. For wind directions in the sector 90°–180°, all levels of power production have a minimum persistence between 80 and 100 min at around 1 pm. Maximum persistence is around midnight varying between 220 min (PUPP) and 140 min (PHPP). It can be seen that for power levels below 50% of maximum production the time of day dependency of persistence is very similar. For power levels above 50% persistence decreases as power production increases. However, the persistence variability with the power level is rather low, for example, maximum persistence at a level of 75% is almost 180 min whereas for a level above 0.05% is 200 min.

For wind directions in the sector 270°–360°, it shows that, for all power levels, the curves for both PUPP and PHPP are similar, i.e. their minima and maxima are located around the same time of the day. For instance, maximal persistence of production is reached at around midday. However, for this direction sector, the higher the power production level, the lower the persistence. For power production above 0.05% of maximum power the persistence is 250 min, persistence of production above 75% of maximum power is only 100 min.

Comparing with the other dominant direction, it can be seen, that they have very different persistence behavior. The maxima and minima are at different times of the day for every power level. The persistence increases as power production decreases, for all power levels in the case of the 270°–360° sector. For the 90°–180° sector it decreases only until 50% of maximum power production. Below that, it remains approximately constant.

7. Conclusions

This paper presents an inhomogeneous Markov process to model wind power production. It is developed using states, which combine information about the wind speed, direction and power variables, using real data recorded by a wind turbine in Portugal. The joint partition of the three-dimensional variable space allows to decrease the number of the model states and, simultaneously, encodes the wind power curve into the Markov model. The transition probabilities are considered to be functions that depend on the time of the day and modeled as Bernstein polynomials. The estimation of the transition matrices is performed by solving a

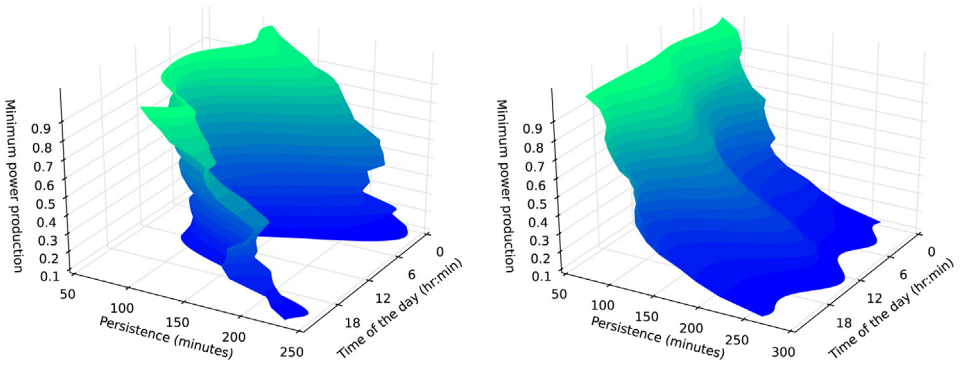


Fig. 8. Time-dependent persistence of power production above $i \cdot 0.05 \cdot p_{max}$, for $i = 1, \ldots, 19$ for direction sectors 90°–180° (left) and 270°–360° (right).

constrained convex optimization problem. Its objective function combines two log-likelihood functions with the purpose to accurately represent both the long-term behavior and the daily fluctuations seen in the original data. To evaluate the statistical properties of the estimated Markov model, synthetic time series are generated and compared with the original data statistics. Results demonstrate that the proposed Markov model can reproduce the diurnal patterns in the data. Moreover it is demonstrated how the persistence of power production throughout the time of the day can be estimated from the Markov process transition matrices.

Acknowledgments

The authors thank the Fundação para a Ciência e a Tecnologia for financial support (SFRH/BD/86934/2012, PEst-OE/MAT/UI0152, FCOMP-01-0124-FEDER-016080 (PTDC/SENENR/1141718/2009)) and GENERG, SA.

References

[1] Anastasiou K, Tsekos C. Persistence statistics of marine environmental parameters from Markov theory, part 1: analysis in discrete time. Appl Ocean Res 1996;18(4):187–99.
[2] Anderson TW, Goodman LA. Statistical inference about Markov chains. Ann Math Stat 1957;28(1):89–110.
[3] Andersson J, Houska B. Towards a computer algebra system with automatic differentiation for use with object-oriented modelling languages. Object-Oriented Modeling Languages; 2010.
[4] Boyd SP, Vandenberghe L. Convex optimization. Cambridge Univ Pr; 2004.
[5] Cancino-Solórzano Y, Gutiérrez-Trashorras AJ, Xiberta-Bernat J. Analytical methods for wind persistence: their application in assessing the best site for a wind farm in the state of Veracruz, Mexico. Renew Energy 2010;35(12):2844–52.
[6] Carapellucci R, Giordano L. The effect of diurnal profile and seasonal wind regime on sizing grid-connected and off-grid wind power plants. Appl Energy 2013;107(0):364–76.
[7] Chen P, Berthelsen K, Bak-Jensen B, Chen Z. Markov model of wind power time series using Bayesian inference of transition matrix. In: Industrial electronics, 2009. IECON '09. 35th Annual conference of IEEE 2009. pp. 627–32.
[8] Corotis RB, Sigl AB, Klein J. Probability models of wind velocity magnitude and persistence. Sol Energy 1978;20(6):483–93.
[9] Kemeny JG, Snell JL. Finite Markov chains. New York: Springer-Verlag; 1976.
[10] Koçak K. A method for determination of wind speed persistence and its application. Energy 2002;27(10):967–73.
[11] Koçak K. Practical ways of evaluating wind speed persistence. Energy 2008;33(1):65–70.
[12] Koçak K. Examination of persistence properties of wind speed records using detrended fluctuation analysis. Energy 2009;34(11):1980–5.
[13] Lopes VV, Scholz T, Estanqueiro A, Novais AQ. On the use of Markov chain models for the analysis of wind power time-series. In: Environment and electrical engineering (EEEIC), 2012 11th international conference on 2012. pp. 770–5.
[14] Masseran N, Razali A, Ibrahim K, Zin WW. Evaluating the wind speed persistence for several wind stations in peninsular Malaysia. Energy 2012;37(1):649–56.
[15] Nfaoui H, Essiarab H, Sayigh A. A stochastic Markov chain model for simulating wind speed time series at Tangiers, Morocco. Renew Energy 2004;29(8):1407–18.
[16] Papaefthymiou G, Klöckl B. MCMC for wind power simulation. Energy Convers IEEE Trans 2008;23(1):234–40.
[17] Pillai S, Suel T, Cha S. The Perron-Frobenius theorem: some of its applications. Signal Process Mag IEEE 2005;22(2):62–75.
[18] Platis A, Limnios N, Le Du M. Dependability analysis of systems modeled by non-homogeneous Markov chains. Reliab Eng Syst Saf 1998;61(3):235–49.
[19] Poje D. Wind persistence in Croatia. Int J Climatol 1992;12(6):569–86.
[20] Pryor S, Barthelmie R. Statistical analysis of flow characteristics in the coastal zone. J Wind Eng Ind Aerodyn 2002;90(3):201–21.
[21] Raischel F, Scholz T, Lopes VV, Lind PG. Uncovering wind turbine properties through two-dimensional stochastic modeling of wind dynamics. Phys Rev E 2013;88(4):42146–57.
[22] Sahin AD, Sen Z. First-order Markov chain approach to wind speed modelling. J Wind Eng Ind Aerodyn 2001;89(34):263–9.
[23] Shamshad A, Bawadi M, Hussin WW, Majid T, Sanusi S. First and second order Markov chain models for synthetic generation of wind speed time series. Energy 2005;30(5):693–708.
[24] Sigl AB, Corotis RB, Won DJ. Run duration analysis of surface wind speeds for wind energy application. J Appl Meteorol 1978;18:156–66.
[25] Suomalainen K, Silva C, ao PF, Connors S. Wind power design in isolated energy systems: impacts of daily wind patterns. Appl Energy 2013;101(0):533–40.
[26] Suomalainen K, Silva C, Ferro P, Connors S. Synthetic wind speed scenarios including diurnal effects: implications for wind power dimensioning. Energy 2012;37(1):41–50.
[27] Wächter A, Biegler LT. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Math Program 2006;106:25–57.
[28] Wen J, Zheng Y, Donghan F. A review on reliability assessment for wind power. Renew Sustain Energy Rev 2009;13(9):2485–94.

Chapter 6

Modeling and analysis of cyclic inhomogeneous Markov processes- a wind turbine case study

The previous chapter described a method to model wind power data as a cyclic time-inhomogeneous Markov process. This chapter revisits the idea of using the discrete Markov model as a data filter for the subsequent analysis in the Langevin framework. As already mentioned in chapter 4, this ensures Markovianity and also overcomes the problem of missing data points. Moreover, the commonly applied estimation procedures for drift and diffusion coefficients require stationarity of the data and are thus not capable of capturing daily variations in the original data, as they are often observed in wind energy production due to the solar cycle.
This chapter presents a method to estimate the drift and diffusion coefficients from the transition matrices computed with the Markov model introduced in Chapter 5. This allows to analyse the dynamical features of the underlying time-dependent stochastic process and monitor their changes throughout the day. The functionality of the method is shown by application to measurements of a wind turbine and comparison to previous findings.

Modeling and analysis of cyclic inhomogeneous Markov processes: a wind turbine case study

Teresa Scholz[1,2], Vitor V. Lopes[3,4], Pedro G. Lind[5], and Frank Raischel[6,7]

[1] Center for Theoretical and Computational Physics, University of Lisbon, Portugal
[2] Energy Analysis and Networks Unit, National Laboratory of Energy and Geology, Lisbon, Portugal
(e-mail: teresa.scholz@lneg.pt)
[3] DEIO-CIO, Science faculty, University of Lisbon, Portugal
[4] Universidad de las Fuerzas Armadas-ESPE, Latacunga, Ecuador
[5] ForWind-Center for Wind Energy Research, Institute of Physics, Carl-von-Ossietzky University of Oldenburg, Oldenburg, Germany
[6] Department of Theoretical Physics, University of Debrecen, Debrecen, Hungary
[7] Center for Geophysics, IDL, University of Lisbon, Portugal

Abstract. A method is proposed to reconstruct a cyclic time-inhomogeneous Markov process from measured data. First, a time-inhomogeneous Markov model is fit to the data, taken here from measurements on a wind turbine. From the time-dependent transition matrices, the time-dependent Kramers-Moyal coefficients of the corresponding stochastic process are computed. Further applications of this method are discussed.
Keywords: time-inhomogeneous Markov process; cyclic Markov process; Kramers-Moyal coefficients.

1 Introduction

Many complex systems can be described, within a certain level of modelization, as stochastic processes. A general stochastic process can be characterized in the linear noise approximation through a Fokker-Planck equation, in continuous variables. For dealing with discrete variables in discrete time steps, often Markov Chains are the models of choice. Although in many cases both approaches converge in the limit of small time steps and increments of the stochastic variables, this correspondence is in general non-trivial[11]. In the Fokker-Planck picture, the so-called Kramers-Moyal (KM) coefficients provide a complete description of a given stochastic process[2].
In the past decades, numerical procedures have been established to estimate the KM coefficients from measured stochastic data, which are applicable for any stationary, i.e. time-homogeneous, Markov process. These methods require large sequences of data, but they are robust[1], have well-known errors and limitations[3], require little intervention and are typically parameter-free[1,5].
However, for non-stationary Markov processes, much fewer methods and results are available to our knowledge. In this case, estimations of the time-dependent KM coefficients can be obtained by two approaches: either the data from the inhomogeneous

3^{rd} *SMTDA Conference Proceedings, 11-14 June 2014, Lisbon Portugal*
C. H. Skiadas (Ed)

process is split into shorter, homogeneous sequences, on which then an estimate of the KM coefficients can be performed through the aforementioned methods[13]. Or, if the inhomogeneous process is also cyclic, a parametrized time-dependent *ansatz* for the KM coefficients can be fit to the data[14]. Compared to the stationary processes, both approaches for the inhomogeneous case require a much higher level of pre-analysis, guesswork and iterative improvement.
In this paper, we present a method that allows to estimate the transition matrices of a time-inhomogeneous Markov model from data. As reported in a previous publication[6], this method provides results with a considerable level of accuracy. Under well-known limitations, the discrete Markov model corresponds to a continuous stochastic process in the form of a Fokker-Planck equation, which is completely characterized, in this case, through its time-dependent KM coefficients. From the transition matrices, we can immediately calculate these KM coefficients, and therefore characterize the dynamical features underlying the time-dependent stochastic process.
We apply this methodology to data from a turbine in a wind park, where measurements of the wind velocity and direction, and the electric power output of the turbine are taken in 10 minute intervals. The results presented from this analysis show the general applicability of our method and are in agreement with previous findings.
This paper is organized as follows. We start in Sec. 2 by introducing both the cyclic time-dependent Markov model and the procedure for extracting stochastic evolution equations directly from data series. In Sec. 3 we describe the data and in Sec. 4 we present the time-dependent functions that define the stochastic evolution of the state of the wind turbine. Section 5 concludes the paper.

2 Methodology

This section describes the methodology used for the data analysis. In Sec. 2.1 the cyclic inhomogeneous Markov model to represent the daily patterns in the data is described and in Sec. 2.2 we explain how stochastic evolution equations are derived directly from the Markov process transition matrices.

2.1 Modelling cyclic time-dependent Markov processes

The goal of this time-inhomogeneous Markov process is to get a model that accurately reproduces the long-term behavior while considering the daily patterns observed in the data. Thus, the proposed objective function combines two maximum likelihood estimators: the first term maximizes the likelihood of the cycle-average probability; and, the second term maximizes the likelihood of the time-dependent probability. The final optimization problem is transformed into a convex one using the negative logarithm of the objective function. This section gives a brief overview over the final optimization problem. A detailed description of the objective function, the parametrization of the time-variant probability functions, and the constraints that must be added to the optimization problem to ensure its Markov properties is provided in [6]. A discrete finite Markov process $\{X_t \in S, t \geq 0\}$ is a stochastic process on a discrete finite state space $S = \{s_1, ..., s_n\}$, $n \in \mathbb{N}$, whose future evolution depends only on its current state [8]. It can be fully described by the conditional probability $Pr\{X_{t+1} = s_j \mid X_t = s_i\}$ of

the Markov process moving to state s_j at time step $t+1$ given that it is in state s_i at time t. It is called the t-th step transition probability, denoted as $p_{i,j}(t)$.
Being time-dependent, the Markov process has associated transition probability matrices P_t that change with time. Considering n possible states, the matrices P_t have dimension $n \times n$ with entries $[P_t]_{i,j} = p_{i,j}(t)$ for all $i, j = 1, \ldots, n$, satisfying $p_{i,j}(t) \geq 0$ and $\sum_j p_{i,j}(t) = 1$.
Markov process is called cyclic with period $T \in \mathbb{R}$, if T is the smallest number, such that $p_{i,j}(mT + r) = p_{i,j}(r)$ for all $m \in \mathbb{N}$ and $0 \leq r < T$. See Ref. [9]. Since this paper deals with discrete data, T and r can considered to be multiples of the time step Δt between successive data points and therefore integers. One can describe the cyclic Markov process by T transition matrices P_r, $r = 0, ..., T-1$. The remainder of time step t modulo T will be denoted as r_t and consequently $r_t = r_{t+mT}$. We fix $T = 1$ day and use $\Delta t = 1$.
In this paper, the transition probabilities $p_{i,j}(z)$ are modeled by Bernstein polynomials, namely

$$p_{i,j}(z) = \sum_{\mu=0}^{k} \beta_{\mu}^{i,j} b_{\mu,k}(z), \tag{1}$$

where $z = r/T$ indicates the time of the day ($T = 1$ day), $b_{\mu,k}(z)$ is the μ-th Bernstein basis polynomial of order k, and $\beta_{\mu}^{i,j} \in \mathbb{R}$. The choice of these polynomials has several advantages properly described in [6].
To maximize the likelihood of the time-dependent transition probabilities given the data, the objective function must consider the time of the day z when the transition happens. The corresponding term of the objective function is thus given by $\sum_{(i,j)_z \in \mathcal{S}_z} \log(p_{i,j}(z))$, where $\mathcal{S}_z$ is the set of observed transitions together with the time z when they happen. This estimator allows to compute the intra-cycle transition probability functions, and thus to represent the daily patterns present in the data.
A second term is added to this function, namely $\sum_{(i,j)\in\mathcal{S}} \log(p_{i,j}^{avg})$, where $\mathcal{S}$ is the set of transitions observed in the data and $p_{i,j}^{avg}$ is the cycle-average (daily) probability of transition from state s_i to s_j. It is given by $p_{i,j}^{avg} = \frac{1}{k+1} \sum_{\mu=0}^{k} \beta_k^{i,j}$. This second term is the maximum likelihood estimator for the daily average probability and its addition to the objective function increases the consistency of the long-term behavior of the Markov process with the data.
Using the resulting overall objective function the optimization problem to be solved for the transition probability coefficients $\beta_{\mu}^{i,j}$ is translated into the minimization of

$$\mathcal{L} = -\sum_{(i,j)\in\mathcal{S}} \log\left(\frac{1}{k+1} \sum_{\mu=0}^{k} \beta_{\mu}^{i,j}\right) - \sum_{(i,j)_z \in \mathcal{S}_z} \log\left(\sum_{\mu=0}^{k} \beta_{\mu}^{i,j} b_{\mu,k}(z)\right) \tag{2}$$

subject to

$$\sum_j \beta_{\mu}^{i,j} = 1 \tag{3a}$$

$$\beta_0^{i,j} = \beta_k^{i,j} = \tfrac{1}{2}(\beta_1^{i,j} + \beta_{k-1}^{i,j}) \tag{3b}$$

$$0 \leq \beta_k^{i,j} \leq 1 \tag{3c}$$

with $i, j = 1, \ldots, n$ and $\mu = 0, \ldots, k$, k being the order of the Bernstein polynomials. Constraint (3a) assures the row-stochasticity of the transition matrices, while the constraint (3b) imposes $\mathcal{C}^0$- and $\mathcal{C}^1$-continuity at $z = 0$. Constraint (3c) bound the transition probabilities between 0 and 1. This constraint is derived using a property of the Bernstein polynomials to always lie in the convex hull defined by their control points $(\frac{k}{\mu}, \beta_\mu)$, $\mu = 0, ..., k$. This convex hull bound can be tightened by subdivision using the de Casteljau algorithm, as described in [6].

2.2 Extracting the stochastic evolution equation

In this section we characterize general stochastic processes through a Fokker-Planck equation. We consider a N-dimensional stochastic process $\mathbf{X} = (X_1(t), \ldots, X_N(t))$ whose probability density function (PDF) $f(\mathbf{X}, t)$ evolves according to the Fokker-Planck equation (FPE) [2]

$$\begin{aligned}\frac{\partial f(\mathbf{X}, t)}{\partial t} = &- \sum_{i=1}^{N} \frac{\partial}{\partial x_i} \left[D_i^{(1)}(\mathbf{X}) f(\mathbf{X}, \mathbf{t}) \right] \\ &+ \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\partial^2}{\partial x_i \partial x_j} \left[D_{ij}^{(2)}(\mathbf{X}) f(\mathbf{X}, t) \right] \quad . \end{aligned} \tag{4}$$

The functions $D_i^{(1)}$ and $D_{ij}^{(2)}$ are the first and second Kramers-Moyal coefficients respectively, more commonly called the drift and diffusion coefficients.
These coefficients provide a complete description of a given stochastic process and are defined as

$$\mathbf{D}^{(k)}(\mathbf{X}) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \frac{\mathbf{M}^{(k)}(\mathbf{X}, \Delta t)}{k!} \quad , \tag{5}$$

where $\mathbf{M}^{(k)}$ are the first ($k = 1$) and second ($k = 2$) conditional moments. $\mathbf{D}^{(1)}$ is the drift vector and $\mathbf{D}^{(2)}$ the diffusion matrix.
If the underlying process is stationary and therefore both drift and diffusion coefficients do not explicitly depend on time t, the conditional moments can be directly derived from the measured data as [1,5]:

$$\begin{aligned} &M_i^{(1)}(\mathbf{X}, \Delta t) = \langle Y_i(t + \Delta t) - Y_i(t) | \mathbf{Y}(t) = \mathbf{X} \rangle \\ &M_{ij}^{(2)}(\mathbf{X}, \Delta t) = \\ &\langle (Y_i(t + \Delta t) - Y_i(t))(Y_j(t + \Delta t) - Y_j(t)) | \mathbf{Y}(t) = \mathbf{X} \rangle \, , \end{aligned} \tag{6}$$

where $\mathbf{Y}(t) = (Y_1(t), \ldots, Y_N(t))$ exhibits the N-dimensional vector of measured variables at time t and $\langle \cdot | \mathbf{Y}(t) = \mathbf{X} \rangle$ symbolizes a conditional averaging over the entire measurement period, where only measurements with $\mathbf{Y}(t) = \mathbf{X}$ are taken into account. In practice binning or kernel based approaches with a certain threshold are applied in order to evaluate the condition $\mathbf{Y}(t) = \mathbf{X}$. See e.g. Ref. [1] for details. If the process is non-stationary and time-inhomogeneous, we must consider an explicit time-dependence of the KM coefficients, which translates into time-dependent conditional moments that can be calculated using a short-time propagator[1]. In our case, this short-time propagator corresponds to the transition probabilities $p_{i,j}(t)$, yielding for the conditional moments

$$M^{(l)}(P_k, v_k, \theta_k, t + \Delta t) = \begin{pmatrix} \sum_j p_{k,j}(t) \left(v_j - v_k\right)^l \\ \sum_j p_{k,j}(t) \left(P_j - P_k\right)^l \\ \sum_j p_{k,j}(t) \left(\theta_j - \theta_k\right)^l \end{pmatrix} . \quad (7)$$

3 Data: wind and power at one wind turbine

The data for this study was obtained from a wind power turbine in a wind park located in a mountainous region in Portugal. The time series consists of a three-year period (2009-2011) of historical data gotten from the turbine data logger. The sampling time of 10 minutes leads to 144 samples each day. The data-set comprises three variables, wind power, speed and direction (nacelle orientation). The wind speed information was collected from the anemometer placed in the wind turbine hub. Due to confidentiality, wind power and speed data values are reported as a fraction of the rated power and the cut-out speed, respectively.
For this Markov model, each state is defined by the values of all three variables, namely the wind speed, wind direction and power output. Figure 1 shows the data observations and the state partitions projected into the wind direction and speed plane (right) and the wind power and speed plane (left). As expected, the observations projected into the wind power and speed plane define the characteristic power curve of the wind turbine.

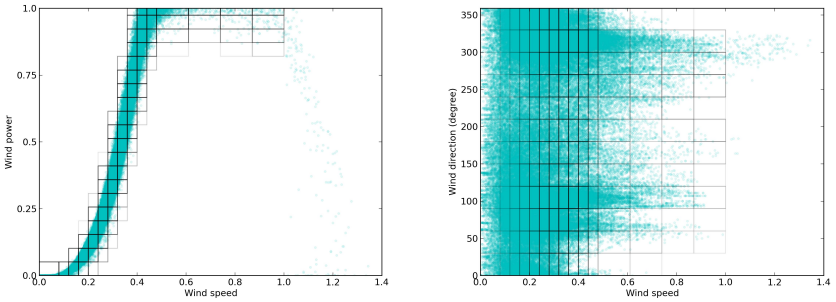


Fig. 1. Representation of all data points projected into the: a) wind direction and speed plane (left); and, b) wind power and speed plane (right). Each rectangle is the projection of a state polyhedron into the two planes. Overall, they define the final state partition for the three-dimensional variable space.

The data space is discretized unevenly to get a good resolution of the high-slope region of the power curve. In a previous work [7], this partition was used in a time-homogeneous Markov chain and proved to lead to an accurate representation of the original data. The wind direction and power are divided by an equally spaced grid leading to 12 and 20 classes, respectively. The wind speed is divided as follows: values below the cut-in speed define one class; between the cut-in and rated wind speed

the discretization is narrowed by selecting 10 classes ; and between the rated and cut-out wind speed discretization is widened and 4 classes are defined. Data points with wind speed above the cut-out wind speed are discarded. The complete state set is constructed by listing all possible combinations of the classes of each variable. Due to physical constraints between the variables, most of the states are empty and can are therefore discarded. This reduces the number of states from 3840 to 778, for this turbine.

To compare the model with the original data, wind power, speed and direction time series were simulated adapting the method described by Sahin and Sen [10] to the cyclic time-inhomogeneous Markov model as follows. First, we compute the cumulative probability transition matrices P_r^{cum} with entries $[P_r^{\text{cum}}]_{i,j} = \sum_{j'=0}^{j} p_{i,j'}(r)$. Then an initial state s_i, i.e. $X_0 = s_i$, is randomly selected. A new datapoint X_{t+1} is generated by uniformly selecting a random number ϵ between zero and one and choosing for X_{t+1} the corresponding state $s_{i'}$ such that the probability of reaching it from the current state s_i fullfils $[P_{r_t}^{\text{cum}}]_{i,i'} \geq \epsilon$.Based on this discrete state sequence, a real value for the wind power/speed/direction variables is generated by sampling each state partition uniformly.

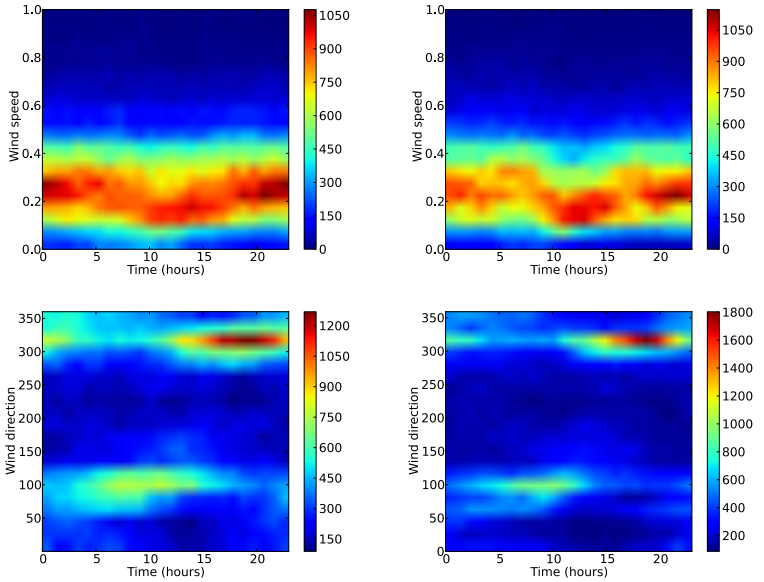


Fig. 2. Two dimensional histograms of the synthetic time-series data, generated with the time-variant Markov model (left) and the original data (right): speed-time (top) and direction-time (bottom).

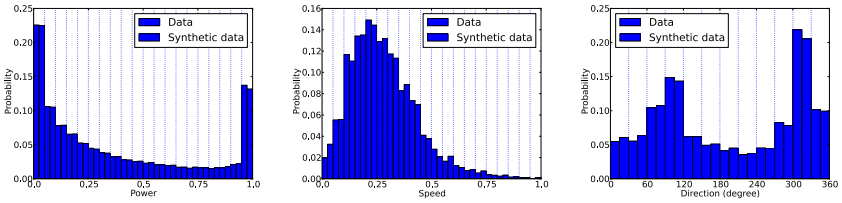


Fig. 3. Comparison of the probability distribution of wind power (left), wind speed (middle) and wind direction (right) of the original with the synthesized data.

Figures 3 and 2 compare the original data with the synthesized data and demonstrate, that the model can capture the data's daily patterns (fig. 2) as well as the long-term statistics (fig. 3).

4 The evolution of drift and diffusion in wind power output

With the procedure outlined in Sec. 2 and having the 144 transition matrices generated as described in Sec. 3 and 2.1, we can now reconstruct the time-dependent stochastic process by calculating the KM coefficients $\mathbf{D}^{(i)}(\hat{\mathbf{X}}, t)$ at each time step $t = 1, \ldots, 144$. Although we obtain the KM coefficients as a function of all three stochastic variables, $[P, v, \theta]$, we here consider only their dependency on the velocity and power production, $\hat{\mathbf{X}} = [\mathbf{P}, \mathbf{v}]$, averaging over the contributions from θ.
The results of this process are presented in Fig.4, where the reconstructed KM coefficients are plotted for three time steps, namely at 6, 12 and 18 hours. The support of the coefficients is limited to the available data which follows the power curves in the v-P plane. From the inspection of Figs. 4, changes in time seem not significative. This means, that even though both the Markov and the stochastic evolution model contain additional degrees of freedom due to their time-dependent formulation, they are capable of capturing the v-P-dependency, which is invariant. This is expected since the wind turbine operation characteristics should not change through the daily cycle. However, it can be seen in fig. 5 that the procedure is capable of detecting even subtle temporal changes in the transition matrix, which lead to strong daily changes in the KM coefficients.
For all plotted times, the drift coefficients $D^{(1)}$ indicate a restoring force towards the power curve, in accordance with previous results[12]. The diffusion coefficients—only the diagonal components are shown here—show an order of magnitude weaker diffusion in the velocity than in the power, where the latter shows a strong component for diffusion in the $P-$ direction for the high slope region of the power curve. These results again are consistent with our previous analysis of a time-homogeneous model [12]. Remarkably, the out-of the $v, P-$ plane diffusion of the direction component $D^{(2)}_{\theta\theta}$ is strongest for both very high and very low velocities, and for intermediate velocities off the power curve.

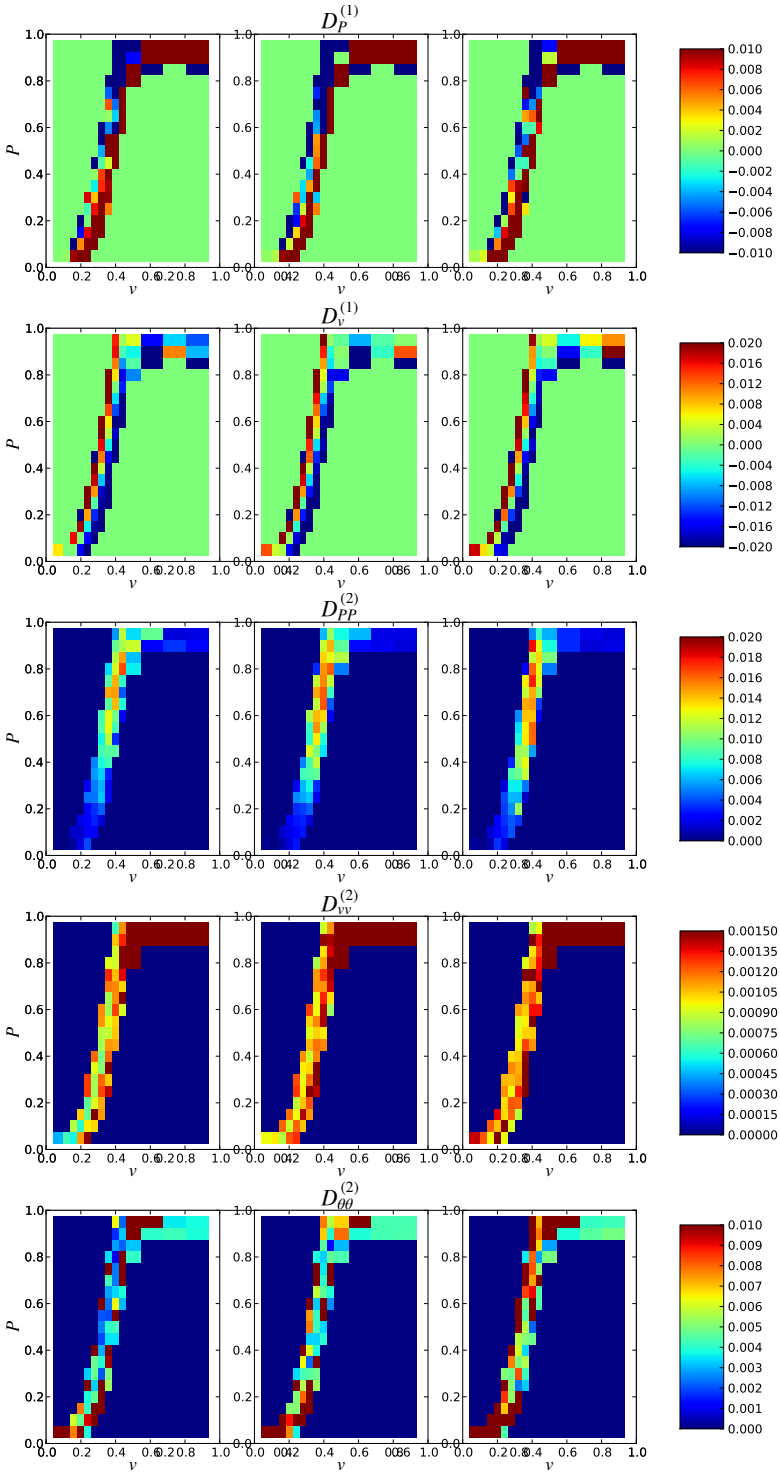


Fig. 4. The first (top two rows) and second (bottom three rows) Kramers-Moyal coefficients for various times (left: 6hours, middle: 12hours, right: 18hours).

Next, we present a closer inspection of the time-dependence of both drift and diffusion, by considering their temporal evolution at a specific point, namely at $(v, P) = (0.34, 0.53)$, which is close to the center of the power curve. Apparently, our method creates smooth curves for the temporal evolution. This is expected since, as a consequence of the parametrization of the Markov model, it can be shown that the conditional moments used to derive the Drift and Diffusion coefficients also can be expressed by Bernstein polynomials in time. Most strikingly, it can be seen that the temporal evolution of both the drift and diffusion coefficient is decoupled between the components. Furthermore, for the same component the evolution of the diffusion coefficient seems to be delayed with respect to the drift. The dominant component is always the power production P, whose drift changes from a positive maximum at 6 h to a negative minimum at 17 h, i.e. the restoring force oscillates from a tendency to higher P values in the morning to a tendency to lower P values in the evening.
It should be noted that the chosen point (v, P) is not necessarily characteristic of the wind field or of the turbine's power production. Other points along the power curve, specifically for low velocities and near the rated wind speed are either more frequent or more characteristic, and their analysis should give increased insight into the temporal evolution of wind speed and power production.

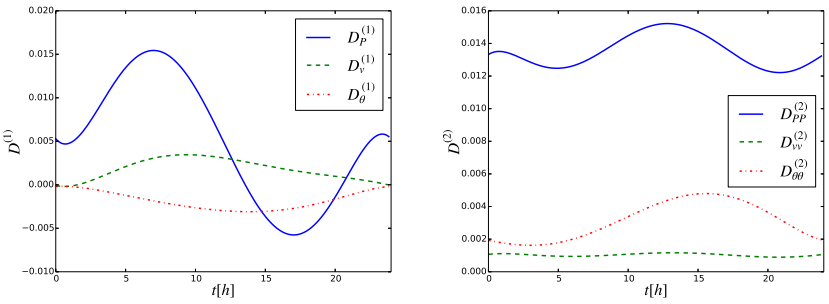


Fig. 5. The first (left) and second (right) Kramers-Moyal coefficient, by components, near the center of the power curve, $(v, P) = (0.34, 0.53)$,as a function of time.

5 Conclusions

We have shown in this paper how a time-dependent multi-dimensional stochastic process can be reconstructed from experimental data. Our method provides results in terms of the time-dependent transition matrix of a Markov model, from which the time-dependent Kramers-Moyal coefficients for a corresponding continuous process can be calculated. Application of this method to data from a turbine in a wind park gives results consistent with a previous time-independent method, and adds surprising new insight into the temporal dynamics of the wind field and the machine power production. Preliminary results have shown that the dependence with time observed in Fig. 5 changes depending which region of the power curve we choose. A more systematic study for the full power-velocity range will be carried out in an extended study.

Future research will also address the question of applicability of our method to more general cases, dealing also with the reliability and relative errors of this approach.

The aforementioned equivalence of the transition Matrix and the KM coefficients is valid if two requirements are fulfilled. First, the transition amplitudes need to have Gaussian shape, which corresponds to the existence of Gaussian noise in the stochastic process. The validity of this assumption has been checked previously for a similar system and can be reasonably assumed in this case. Secondly, the binning of the stochastic variables for determining the Markov process transition matrix must be small enough[11]. We will investigate the validity of this assumption and the corresponding errors in a forthcoming publication.

Acknowledgements

The authors acknowledge helpful discussions with David Kleinhans, who provided the basic idea to calculate the Kramers-Moyal coefficients directly. The authors thank Fundação para a Ciência e a Tecnologia for financial support under PEst-OE/FIS/UI0618/2011, PEst-OE/MAT/UI0152/2011,
FCOMP-01-0124-FEDER-016080 and SFRH/BD/86934/2012 (TS). VVL thanks the Prometeo Project of SENESCYT (Ecuador) for financial support. PGL thanks the German Environment Ministry for financial support (0325577B). This work is part of a bilateral cooperation DRI/DAAD/1208/2013 supported by FCT and Deutscher Akademischer Auslandsdienst (DAAD). FR assisted in fundamental research in the frame of TÁMOP 4.2.4. A/2-11-1-2012-0001 National Excellence Program ? Elaborating and operating an inland student and researcher personal support system, was realised with personal support. The project was subsidized by the European Union and co-financed by the European Social Fund. FR would like to thank F.Kun, Univ. Debrecen, for his hospitality.

References

1. R. Friedrich, J. Peinke, M. Sahimi and M.R.R. Tabar, Phys. Rep. **506** 87 (2011).
2. H. Risken, *The Fokker-Planck Equation* (Springer, Heidelberg,1984).
3. D. Kleinhans, Phys. Rev. E **85** 026705 (2012).
4. F. Boettcher, J. Peinke, D. Kleinhans, R. Friedrich, P.G. Lind, M. Haase, Phys. Rev. Lett. **97** 090603 (2006).
5. P.G. Lind, M. Haase, F. Boettcher, J. Peinke, D. Kleinhans and R. Friedrich, Physical Review E 81 041125 (2010).
6. T. Scholz, V.V. Lopes, A. Estanqueiro, "A cyclic time-dependent markov process to model daily patterns in wind turbine power production", Energy **67**(0), 557-568, 2014.
7. V.V. Lopes, T. Scholz, A. Estanqueiro, and A.Q. Novais, "On the use of Markov chain models for the analysis of wind power time series", Environment and Electrical Engineering (EEEIC), 2012 11th International Conference on, 770-775, 2012.
8. J.G. Kemeny and J.L. Snell, "Finite Markov Chains", New York : Springer-Verlag, 1976.
9. A. Platis and N. Limnios and M. Le Du, "Dependability analysis of systems modeled by non-homogeneous Markov chains", Reliability Engineering and System Safety, **61**(3), 235-249, 1998.
10. A.D. Sahin and Z. Sen, "First-order Markov chain approach to wind speed modelling", Journal of Wind Engineering and Industrial Aerodynamics, **89**(3-4), 263-269, 2001.

11. N.G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North Holland, 2007).
12. F.Raischel, T.Scholz, V.V.Lopes and P.G.Lind, "Uncovering wind turbine properties through two-dimensional stochastic modeling of wind dynamics", Phys. Rev. E **88** 042146 (2013).
13. A. M.van Mourik, A.Daffertshofer, P.J.Beek, "Estimating Kramers Moyal coefficients in short and non-stationary data sets", Physics Letters A **351**, Issue 1-2, 13-17, 2006.
14. C.Micheletti, G.Bussi, A.Laio, "Optimal Langevin Modeling of out-of-Equilibrium Molecular Dynamics Simulations", The Journal of Chemical Physics **129**, 7, 2008.

Chapter 7

Parameter free resolution of the superposition of stochastic signals

The previous chapters illustrated the application of the Langevin method to data from wind energy production. Several common challenges in the modeling of measurement data were addressed, such as ensuring stationarity and Markovianity of the data, overcoming difficulties arising from missing data points and short time-series and also dealing with the usually available 10-minute average data. Yet another challenge when dealing with real world data are the measurements themselves. Often, the available time-series are subject to noise that is associated to the measurement devices, which spoils the data by hiding the underlying stochastic process. When the Langevin method is applied to this spoiled data, the estimated moments are also spoiled which in return yields distorted drift and diffusion coefficients that cannot grasp the dynamics of the underlying process. To illustrate this, for the example discussed in this chapter, the first three joint (top) and noisy moments (bottom) are shown in figure 7.1.

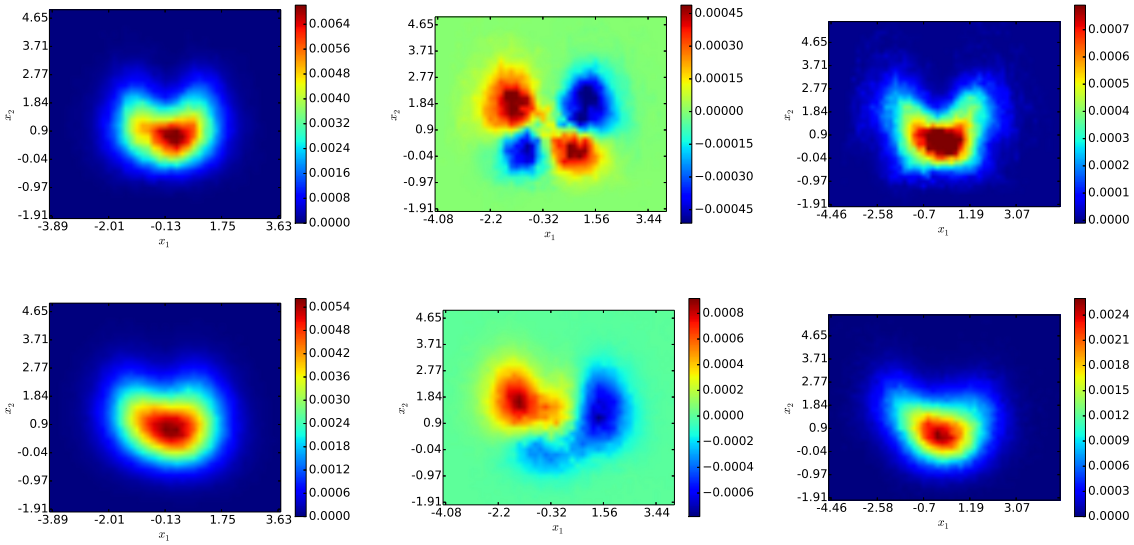


Figure 7.1: Zeroth (left), first (middle) and second (right) empirical (top) and noisy joint moments (bottom) of the example presented in the following of this chapter.

This chapter presents a method to compute the drift and diffusion coefficients in the presence of measurement noise[a]. This has already been done by other authors using a different *ansatz*, namely assuming a certain functional form of the Langevin parameters. The method described in this paper is parameter-free and extracts all coefficients directly from the data.

[a]The corresponding paper is under submission.

Physics Letters A ••• (••••) •••–•••

Contents lists available at ScienceDirect

Physics Letters A

www.elsevier.com/locate/pla

Parameter-free resolution of the superposition of stochastic signals

Teresa Scholz [a], Frank Raischel [b,c], Vitor V. Lopes [d,e], Bernd Lehle [f], Matthias Wächter [f], Joachim Peinke [f], Pedro G. Lind [f,g]

[a] Center for Theoretical and Computational Physics, University of Lisbon, Portugal
[b] Center for Geophysics, IDL, University of Lisbon, Portugal
[c] Closer Consulting, Av. Eng. Duarte Pacheco Torre 1 15^0, 1070-101 Lisboa, Portugal
[d] DEIO-CIO, University of Lisbon, Portugal
[e] UTEC–Universidad de Ingeniería y Tecnología, Lima, Peru
[f] Institute of Physics and ForWind, Carl-von-Ossietzky University of Oldenburg, Oldenburg, Germany
[g] Institute of Physics, University of Osnabrück, Osnabrück, Germany

ARTICLE INFO

Article history:
Received 30 August 2016
Accepted 25 September 2016
Available online xxxx
Communicated by F. Porcelli

Keywords:
Stochastic processes
Measurement noise
Observational noise
Signal reconstruction
Signals superposition

ABSTRACT

This paper presents a direct method to obtain the deterministic and stochastic contribution of the sum of two independent stochastic processes, one of which is an Ornstein–Uhlenbeck process and the other a general (non-linear) Langevin process. The method is able to distinguish between the stochastic processes, retrieving their corresponding stochastic evolution equations. This framework is based on a recent approach for the analysis of multidimensional Langevin-type stochastic processes in the presence of strong measurement (or observational) noise, which is here extended to impose neither constraints nor parameters and extract all coefficients directly from the empirical data sets. Using synthetic data, it is shown that the method yields satisfactory results.

1. Introduction and motivation

An important topic in the analysis of time-series of complex dynamical systems is the extraction of the underlying process dynamics. Often it is possible to reveal the deterministic and stochastic contributions of the underlying stochastic process using the Itô–Langevin equation, a stochastic equation that describes the evolution of a stochastic variable. The deterministic and stochastic contributions are given by the so-called drift and diffusion coefficients, which can be directly derived from data via joint moments [4]. This approach has been applied successfully to several areas [5], for example the description of turbulence [4,17], the analysis of climate data [10], financial data [16], biological systems [19] and wind energy production [15,13].

However, typically the time-series to be analyzed is subject to noise, which is associated to the measurement devices or other sources. This so-called measurement noise, also known as observational noise, is not involved in the dynamics of the original signal. Nevertheless it spoils the data series by hiding the underlying stochastic process. In this case, the joint moments are not accessible but only their "noisy" analogues. Several approaches have been published to overcome this challenge. The authors of Refs. [3] and [9] introduced a method that allows the estimation of the drift and diffusion coefficients in the presence of strong, delta-correlated Gaussian measurement noise. An alternative approach was presented by Lehle [7] that can deal with strong, exponentially correlated Gaussian noise in one dimension, which was extended to be applicable to multidimensional time-series [8]. This approach is the basis of the method presented in this paper. Here, instead of using a parameterized form of the coefficients defining the stochastic processes, the method extracts all coefficients directly from the data.

In a more general framework, the paper presents a method which allows to distinguish between two superposed signals, i.e. extract their respective evolution equations, if one of them is an Ornstein–Uhlenbeck process. Specifically, the method serves to extract the measurement noise parameters as well as the drift and diffusion coefficients describing the stochastic process from the original data, which henceforth are called "noisy" data. This allows to separate the two stochastic signals: the measurement noise, described by an Ornstein–Uhlenbeck process, and the underlying general Langevin process. The method can be applied to a set of N coupled stochastic variables superposed with a set of N sources of correlated measurement noise and the code is accessible by request to the authors.

E-mail address: tascholz@fc.ul.pt (T. Scholz).

http://dx.doi.org/10.1016/j.physleta.2016.09.057
0375-9601/

The paper is structured as follows. The theoretical background of the Langevin analysis of stochastic processes and the extraction of the coefficients from data is briefly summarized in Sec. 2. Section 3 gives an overview of the method to obtain those coefficients in the presence of measurement noise. Subsequently, the two main challenges in the method are presented: a) the solution of a nonlinear equation system to obtain the measurement noise parameters, which is described in Sec. 4, and b) the solution of a system of convolution equations to estimate the joint moments of the underlying stochastic process, which is described in Sec. 5. The results of application to a synthetic data set are shown in Sec. 6, demonstrating the accuracy of the presented approach as well as its limits. Section 7 discusses possible applications of the method and concludes the paper.

2. A general model for noisy stochastic processes

The evolution of a stochastic variable can be described by the Itô–Langevin equation, a stochastic equation defined by a deterministic contribution (drift) and fluctuations from possible stochastic sources (diffusion). For the general case of a N-dimensional stochastic process $\mathbf{X}(t)$ the equation is given by:

$$d\mathbf{x} = \mathbf{D}^{(1)}(\mathbf{x})dt + \sqrt{\mathbf{D}^{(2)}(\mathbf{x})}d\mathbf{W}(t), \tag{1}$$

where $d\mathbf{W}$ denotes a vector of increments of independent Wiener processes with $\langle d\mathbf{W}_i\rangle = 0$ and $\langle d\mathbf{W}_i, d\mathbf{W}_j\rangle = \delta_{ij}dt\ \forall i, j = 1,\dots,N$, where $\langle\rangle$ denotes the average and δ_{ij} the Kronecker delta. Functions $\mathbf{D}^{(1)}(\mathbf{x})$ and $\mathbf{D}^{(2)}(\mathbf{x})$ are the Kramers–Moyal coefficients of the corresponding Fokker–Planck equation that describes the evolution of the conditional probability density function. In the case the distribution of initial conditions is known one can derive the evolution equation of the joint probability density function $f(\mathbf{x}, t)$ of the stochastic variables $\mathbf{x}$. It is given by:

$$\frac{\partial f(\mathbf{x},t)}{\partial t} = -\sum_{i=1}^{N}\frac{\partial}{\partial x_i}\left[D_i^{(1)}(\mathbf{x})f(\mathbf{x},t)\right] + \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\partial^2}{\partial x_i \partial x_j}\left[D_{ij}^{(2)}(\mathbf{x})f(\mathbf{x},t)\right]. \tag{2}$$

The Kramers–Moyal coefficients, also called the drift ($\mathbf{D}^{(1)}(\mathbf{x})$) and diffusion ($\mathbf{D}^{(2)}(\mathbf{x})$) coefficients, can be directly derived from measurements [5]. However, here we consider that each measured stochastic variable is the sum of two independent stochastic processes X and Y:

$$\mathbf{X}^*(t) = \mathbf{X}(t) + \mathbf{Y}(t). \tag{3}$$

Since such a situation can be regarded as having a set of N stochastic signals $\mathbf{X}(t)$ spoiled by a set of N sources of measurement noise $\mathbf{Y}(t)$, we call the variables $\mathbf{X}^*(t)$ a N-dimensional *noisy* stochastic process. Fig. 1 shows a specific example of such superposition of stochastic processes that will be addressed below in detail, plotting the first component of $\mathbf{X}^*$, $\mathbf{X}$ and $\mathbf{Y}$. We assume the measurement noise $\mathbf{Y}(t)$ to be described by an Ornstein–Uhlenbeck process in N dimensions:

$$d\mathbf{y}(t) = -\mathbf{A}\mathbf{y}(t)dt + \sqrt{\mathbf{B}}d\mathbf{W}(t), \tag{4}$$

where $\mathbf{A}$ and $\mathbf{B}$ are $N \times N$ matrices, $\mathbf{B}$ is symmetric positive semi-definite and the eigenvalues of $\mathbf{A}$ have a positive real part. Thus, the N-dimensional noisy stochastic process $\mathbf{X}^*$ is modeled by Eqs. (3) and (4) together. Note that here and throughout the paper $\mathbf{x}$ denotes the accessible values of any of the involved stochastic processes $\mathbf{X}(t)$, $\mathbf{X}^*(t)$ or $\mathbf{Y}(t)$.

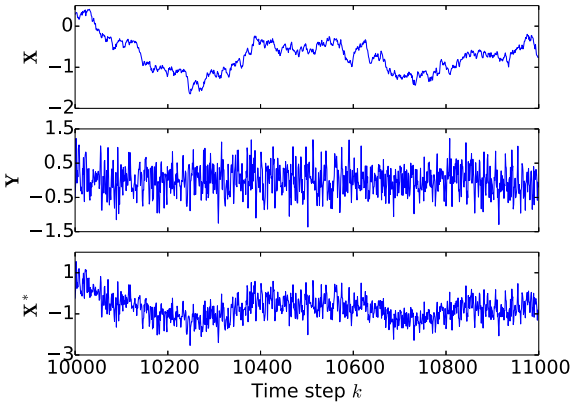


Fig. 1. Illustration of a stochastic process $\mathbf{X}(t)$ (top), governed by a nonlinear Langevin equation (Eq. (1)), a correlated measurement noise $\mathbf{Y}(t)$ (middle), governed by a Ornstein–Uhlenbeck process (Eq. (4)) and the superposition of both processes $\mathbf{X}^*(t) = \mathbf{X}(t) + \mathbf{Y}(t)$ (bottom).

3. From data to model: the inverse problem

This section explains how to obtain the drift and diffusion coefficients along with the measurement noise parameters from noisy data $\mathbf{X}^*(t)$. The methodology is sketched in Fig. 2 and the idea behind it is that, if the measurement noise is independent of the stochastic process, it is possible to derive an equation system that relates the noisy moments $m^{*(0)}(\mathbf{x})$, $\mathbf{m}^{*(1)}(\mathbf{x},\tau)$ and $\mathbf{m}^{*(2)}(\mathbf{x},\tau)$ with the measurement noise-free moments $m^{(0)}(\mathbf{x})$, $\mathbf{m}^{(1)}(\mathbf{x},\tau)$ and $\mathbf{m}^{(2)}(\mathbf{x},\tau)$ and solving it in a parameter-free way is the heart of this paper.

The system of equations (for a derivation see Appendix A) is given by:

$$m^{*(0)}(\mathbf{x}) \equiv \int_{\mathbf{x}'}\rho^*(\mathbf{x},\mathbf{x}',\tau)d\mathbf{x}' = \rho_Y(\mathbf{x}) * m^{(0)}(\mathbf{x}), \tag{5a}$$

$$m_i^{*(1)}(\mathbf{x},\tau) \equiv \int_{\mathbf{x}'}(x_i'(t+\tau) - x_i(t))\rho^*(\mathbf{x},\mathbf{x}',\tau)d\mathbf{x}' = \rho_Y(\mathbf{x}) * m_i^{(1)}(\mathbf{x},\tau) + H_i^{(1)}(\mathbf{A},\mathbf{B},m^{*(0)}(\mathbf{x})), \tag{5b}$$

$$m_{ij}^{*(2)}(\mathbf{x},\tau) \equiv \int_{\mathbf{x}'}(x_i'(t+\tau) - x_i(t))(x_j'(t+\tau) - x_j(t))\rho^*(\mathbf{x},\mathbf{x}',\tau)d\mathbf{x}' = \rho_Y(\mathbf{x}) * m_{ij}^{(2)}(\mathbf{x},\tau) + H_{ij}^{(2)}(\mathbf{A},\mathbf{B},m^{*(0)}(\mathbf{x}),\mathbf{m}^{*(1)}(\mathbf{x},\tau)), \tag{5c}$$

where $i, j = 1,\dots,N$ and

$$\rho^*(\mathbf{x}) = f(\mathbf{x},t), \tag{6a}$$

$$\rho^*(\mathbf{x},\mathbf{x}',\tau) = f(\mathbf{x},t;\mathbf{x}',t+\tau), \tag{6b}$$

are the one and two-point probability density functions of the noisy data, respectively, and

$$\rho_Y(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N|\det(\mathbf{V})|}}e^{-\frac{1}{2}\mathbf{x}^T\mathbf{V}^{-1}\mathbf{x}}, \tag{7}$$

is the probability density function of the measurement noise, assuming that it is distributed with a normalized Gauss function $\mathcal{G}(\mathbf{x},0,\mathbf{V})$ with zero average and covariance $\mathbf{V}$. The functions $H_i^{(1)}(\mathbf{A},\mathbf{B},m^{*(0)}(\mathbf{x}))$ and $H_{ij}^{(2)}(\mathbf{A},\mathbf{B},m^{*(0)}(\mathbf{x}),\mathbf{m}^{*(1)}(\mathbf{x},\tau))$ are given by

Please cite this article in press as: T. Scholz et al., Parameter-free resolution of the superposition of stochastic signals, Phys. Lett. A (2016), http://dx.doi.org/10.1016/j.physleta.2016.09.057

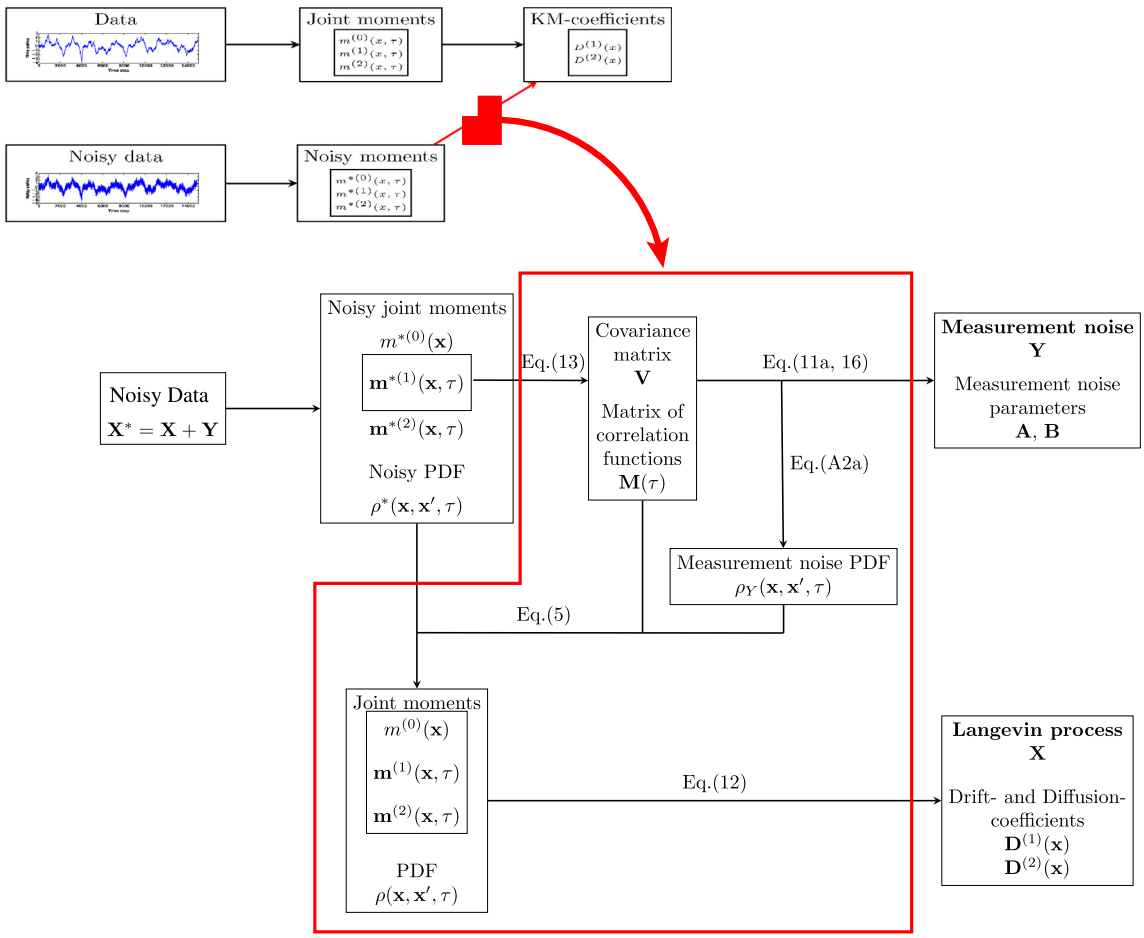


Fig. 2. Graphic representation of the methodology to obtain drift and diffusion coefficients from noisy data. Functional arguments are omitted for better readability.

$$H_i^{(1)}(\mathbf{A},\mathbf{B},m^{*(0)}(\mathbf{x})) = \sum_{k=1}^{N} Q_{ik}(\tau)\frac{\partial}{\partial x_k} m^{*(0)}(\mathbf{x}) \tag{8}$$

$$\begin{aligned} H_{ij}^{(2)}&(\mathbf{A},\mathbf{B},m^{*(0)}(\mathbf{x}),\mathbf{m}^{*(1)}(\mathbf{x},\tau)) \\ &= \left(Q_{ij}(\tau) + Q_{ji}(\tau) - \sum_{k=1}^{N}\sum_{l=1}^{N} Q_{ik}(\tau)Q_{jl}(\tau)\frac{\partial}{\partial x_k}\frac{\partial}{\partial x_l}\right) m^{*(0)}(\mathbf{x}) \\ &\quad + \sum_{k=1}^{N} Q_{ik}(\tau)\frac{\partial}{\partial x_k} m_j^{*(1)}(\mathbf{x},\tau) + \sum_{l=1}^{N} Q_{jl}(\tau)\frac{\partial}{\partial x_l} m_i^{*(1)}(\mathbf{x},\tau), \end{aligned} \tag{9}$$

where the elements $Q_{ij}(\tau)$ define the $N \times N$-matrix

$$\mathbf{Q}(\tau) = \mathbf{Q}(k\Delta t) = (\mathbf{Id} - \mathbf{M}(k\Delta t))\mathbf{V}, \tag{10}$$

with $\mathbf{Id}$ the identity matrix. $\mathbf{V}$ is the covariance matrix and $\mathbf{M} = \mathbf{M}(\Delta t)$ the matrix of decaying correlation functions of the measurement noise. Both are related to the measurement noise parameters $\mathbf{A}$ and $\mathbf{B}$ (see Eq. (4)) through

$$\mathbf{M}(k\Delta t) = e^{-\mathbf{A}k\Delta t} \tag{11a}$$

$$\mathbf{V} = \int_0^\infty e^{-\mathbf{A}s}\mathbf{B}e^{-\mathbf{A}^T s}ds, \tag{11b}$$

where $\mathbf{A}^T$ denotes the transpose of matrix $\mathbf{A}$. Throughout the paper, τ refers to the time-lag, also expressed as a multiple of the discretization time-step Δt, i.e. $\tau = k\Delta t$ for some integer k.

Solving Eqs. (5) is done in three steps. In a first step, the noisy probability density function $\rho^*(\mathbf{x},\mathbf{x}',\tau)$ as well as the zeroth $m^{*(0)}(\mathbf{x})$, first $\mathbf{m}^{*(1)}(\mathbf{x},\tau)$ and second $\mathbf{m}^{*(2)}(\mathbf{x},\tau)$ noisy joint moments are extracted from the data.

In a second step, using the first noisy joint moments $\mathbf{m}^{*(1)}(\mathbf{x},\tau)$ (Eq. (5b)) the noise source is characterized by deriving its parameter matrices $\mathbf{A}$ and $\mathbf{B}$ through the solution of a nonlinear equation system. Its construction and solution are presented in Sec. 4.

In a third step, using the obtained noise parameters as well as the noisy moments, Eqs. (5) can be solved, as fully described in Sec. 5. After this three-step procedure, the drift and diffusion coefficients are computed from the joint moments $m^{(0)}(\mathbf{x})$, $\mathbf{m}^{(1)}(\mathbf{x},\tau)$ and $\mathbf{m}^{(2)}(\mathbf{x},\tau)$ as

$$D_i^{(1)}(\mathbf{x}) = \lim_{\tau\to 0}\frac{1}{\tau}\frac{m_i^{(1)}(\mathbf{x},\tau)}{m^{(0)}(\mathbf{x})}, \tag{12a}$$

$$D_{ij}^{(2)}(\mathbf{x}) = \lim_{\tau\to 0}\frac{1}{\tau}\frac{m_{ij}^{(2)}(\mathbf{x},\tau)}{m^{(0)}(\mathbf{x})}, \tag{12b}$$

for $i, j = 1,\ldots,N$.

4. Extracting the measurement noise parameter matrices

The measurement noise parameters $\mathbf{A}$ and $\mathbf{B}$ (see Eq. (4)) can be obtained by solving a nonlinear equation system. Here, a solution is approximated in the least square sense by solving an optimization problem, which is formulated making use of the symbolic framework for algorithmic differentiation and numeric optimization CasADi [2]. The optimization is performed by a nonlinear interior-point solver called Ipopt [18]. The extensive discussion of the optimization problem, i.e. its objective function and constraints is the subject of this section.

Please cite this article in press as: T. Scholz et al., Parameter-free resolution of the superposition of stochastic signals, Phys. Lett. A (2016), http://dx.doi.org/10.1016/j.physleta.2016.09.057

4.1. The equation system

To compute the two matrices $\mathbf{A}$ and $\mathbf{B}$ defining the measurement noise (see Eq. (4)) consider Eq. (5b). Multiplication with x_j and integration over $\mathbf{x}$ yields

$$\int_{\mathbf{x}} m_i^{*(1)}(\mathbf{x},\tau)x_j d\mathbf{x} = \int_{\mathbf{x}} x_j \rho_Y(\mathbf{x}) * m_i^{(1)}(\mathbf{x},\tau) d\mathbf{x} + \int_{\mathbf{x}} x_j H_i^{(1)}(\mathbf{A},\mathbf{B},m^{*(0)}(\mathbf{x}))d\mathbf{x}, \tag{13}$$

where $i, j = 1,\dots,N$.

The integral on the left hand-side of Eq. (13) can be directly computed from the data. The first integral on the right hand-side is a function of $\tau = k\Delta t$ alone and can thus be approximated by a polynomial in τ:

$$\int_{\mathbf{x}} x_j \rho_Y(\mathbf{x}) * m_i^{(1)}(\mathbf{x},\tau)d\mathbf{x} = \sum_{\nu=1}^{\nu_{max}} P_{ij}^{(\nu)}(k\Delta t)^\nu. \tag{14}$$

The second integral is given by (see Appendix B and Eq. 10)

$$\int_{\mathbf{x}} x_j H_i^{(1)}(\mathbf{A},\mathbf{B},m^{*(0)}(\mathbf{x}))d\mathbf{x} = -\sum_{l=1}^{N}(\delta_{il} - (M(k\delta t))_{il})V_{lj} = -Q_{ij}(\tau), \tag{15}$$

where $k \in \mathcal{K} \subset \mathbb{N}$, the set of time increments with cardinality k_{max} used for the estimation of the noisy moments from the data.

The unknowns of Eq. (13) are the coefficients $\mathbf{P}^{(\nu_1)},\dots,\mathbf{P}^{(\nu_{max})}$, auxiliary $N \times N$-matrices, the matrix $\mathbf{M} = \mathbf{M}(\Delta t)$, from which $\mathbf{A}$ can be derived as the matrix logarithm of $\mathbf{M}$ (Eq. (11a)), and the covariance matrix $\mathbf{V}$ of the measurement noise, from which $\mathbf{B}$ can be obtained through (see Appendix C)

$$\mathbf{B} = \mathbf{V}\mathbf{A}^T + \mathbf{A}\mathbf{V}. \tag{16}$$

4.2. The objective function

To obtain a numerical estimation of the measurement noise parameters, Eqs. (13) is solved in the least square sense and the summed squares of the differences between its left and right hand side is minimized. This objective function is formulated making use of a lifting approach [1], i.e. additional variables $\Omega^{(k)}$, $k \in \mathcal{K}$, and therefore additional degrees of freedom are introduced.

Using the notation $\mathbf{M} = \mathbf{M}(\Delta t)$ and Eq. (11a) the formulation of the objective function F is given by

$$F = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k\in\mathcal{K}}\left[Z_{ij}(k\Delta t) - \sum_{\nu=1}^{\nu_{max}} P_{ij}^{(\nu)}(k\Delta t)^\nu - V_{ij} + \sum_{l=1}^{N}\Omega_{il}^{(k)}V_{lj}\right]^2, \tag{17}$$

where $Z_{ij}(k\Delta t) = Z_{ij}(\tau)$ is the left-hand side of Eq. (13), which can be computed directly from the data sets, and the k_{max} constraints

$$\Omega^{(1)} - \mathbf{M} = 0 \tag{18a}$$

$$\Omega^{(k_{l+1})} - \Omega^{(k_l)}\mathbf{M}^{k_{l+1}-k_l} = 0, \tag{18b}$$

where $l = 1,\dots,k_{max}-1$, are added to the optimization problem. For sets $\mathcal{K}$ of consecutive integers, i.e. $\mathcal{K} = \{1,\dots,k_{max}\}$, this formulation eliminates the powers of $\mathbf{M}$. For sets $\mathcal{K} = \kappa\{1,\dots,k_{max}\} = \{\kappa,\dots,\kappa k_{max}\}$, $\kappa \in \mathbb{N}$, the powers of $\mathbf{M}$ can also be eliminated by minimizing the objective function subject to constraint (18b) for $\mathbf{M}' = \mathbf{M}^\kappa$ and subsequently computing the κ-th root of $\mathbf{M}'$ using the Eigenvalue decomposition. Note that an optimal solution $\mathbf{M}'$ of Eq. (13) subject to Eq. (18b) might have negative or complex Eigenvalues. In this case it is necessary to choose a different set of time-steps $\mathcal{K}$ and/or a different polynomial order ν_{max} and repeat the computations.

4.3. Constraints

The minimization of the objective function F has to be performed with respect to constraints that ensure that $\mathbf{V}$ is symmetric and positive definite and $\mathbf{M}$ is stable. The constraint ensuring symmetry and positive definiteness of the matrix $\mathbf{V}$ is formulated employing the Cholesky decomposition, which is a decomposition based on a $N \times N$-lower triangular matrix with strictly positive diagonal entries. For positive definite matrices the Cholesky decomposition is unique and therefore, symmetry and positive definiteness of $\mathbf{V}$ can be imposed by

$$\mathbf{V} = \mathbf{L}_V\mathbf{L}_V^T, \tag{19}$$

where, using an exponential *Ansatz*, $\mathbf{L}_V$ is given as a lower triangular matrix with positive diagonal elements by

$$\mathbf{L}_V(i,j) = \begin{cases} 0 & \text{if } j > i \\ e^{v_{ij}} & \text{if } j = i \\ v_{ij} & \text{if } j < i, \end{cases} \tag{20}$$

with $v_{ij} \in \mathbb{R}$.

The matrix $\mathbf{M}(k\Delta t)$ is exponentially decaying with k [Eq. (11a)], which holds if and only if $\mathbf{M}$ is stable, i.e. all its eigenvalues λ_i are in the unit circle of the complex plane, $|\lambda_i| < 1$, $\forall i = 1,\dots,N$. A theorem from stability theory [12] states, that if there are two symmetric positive definite matrices $\mathbf{U}$ and $\mathbf{C}$ satisfying

$$\mathbf{U} - \mathbf{M}^T\mathbf{U}\mathbf{M} = \mathbf{C}, \tag{21}$$

then the matrix $\mathbf{M}$ is stable. Therefore, for $\mathbf{M}$ to be stable, $\mathbf{U} - \mathbf{M}^T\mathbf{U}\mathbf{M}$ needs to be symmetric positive definite. Using the Schur complement [21] yields

$$\begin{pmatrix} \mathbf{U} & \mathbf{M}^T\mathbf{U} \\ \mathbf{U}\mathbf{M} & \mathbf{W} \end{pmatrix} \tag{22}$$

as a symmetric positive definite matrix. Thus, again using the Cholesky decomposition, the stability of $\mathbf{M}$ can be formulated as

$$\begin{pmatrix} \mathbf{U} & \mathbf{M}^T\mathbf{U} \\ \mathbf{U}\mathbf{M} & \mathbf{U} \end{pmatrix} = \begin{pmatrix} \mathbf{L}_U & 0 \\ \mathbf{E} & \mathbf{L}_C \end{pmatrix}\begin{pmatrix} \mathbf{L}_U^T & \mathbf{E}^T \\ 0 & \mathbf{L}_C^T \end{pmatrix}, \tag{23}$$

where $\mathbf{L}_U$, $\mathbf{L}_C$ are lower $N \times N$-triangular matrices with positive diagonal elements and $\mathbf{E}$ is a full-rank $N \times N$-matrix. Equation (23) leads to additional constraints, namely:

$$\mathbf{U} = \mathbf{L}_U\mathbf{L}_U^T, \tag{24a}$$

$$\mathbf{M}^T\mathbf{U} = \mathbf{L}_U\mathbf{E}^T, \tag{24b}$$

$$\mathbf{U} = \mathbf{E}\mathbf{E}^T + \mathbf{L}_C\mathbf{L}_C^T, \tag{24c}$$

where

$$\mathbf{L}_C(i,j) = \begin{cases} 0 & \text{if } j > i, \\ e^{c_{ij}} & \text{if } j = i, \\ c_{ij} & \text{if } j < i, \end{cases} \tag{25}$$

$$\mathbf{L}_U(i,j) = \begin{cases} 0 & \text{if } j > i, \\ e^{u_{ij}} & \text{if } j = i, \\ u_{ij} & \text{if } j < i, \end{cases} \tag{26}$$

with $c_{ij}, u_{ij} \in \mathbb{R}$.

Please cite this article in press as: T. Scholz et al., Parameter-free resolution of the superposition of stochastic signals, Phys. Lett. A (2016), http://dx.doi.org/10.1016/j.physleta.2016.09.057

Thus, the full optimization problem is given by the minimization of F (see Eq. (17)) subject to Eqs. (18b), (19) and (24).

5. Obtaining the joint moments

The previous section describes the numerical solution of Eqs. (13), which is needed to obtain the conditional moments $m^{(0)}(\mathbf{x})$, $\mathbf{m}^{(1)}(\mathbf{x},\tau)$ and $\mathbf{m}^{(2)}(\mathbf{x},\tau)$. Namely, from the matrices $\mathbf{M}$ and $\mathbf{V}$ the functions $H_i^{(1)}(\mathbf{A},\mathbf{B},m^{*(0)}(\mathbf{x}))$ and $H_{ij}^{(2)}(\mathbf{A},\mathbf{B},m^{*(0)}(\mathbf{x}),\mathbf{m}^{*(1)}(\mathbf{x},\tau))$ are computed and introduced to Eqs. (5). Solving this equation system yields the measurement noise-free joint moments $m^{(0)}(\mathbf{x})$, $\mathbf{m}^{(1)}(\mathbf{x},\tau)$ and $\mathbf{m}^{(2)}(\mathbf{x},\tau)$. Again, this is done in the least square sense and the formulation of the corresponding optimization problems is presented in this section.

5.1. The objective functions

Equations (5) are solved sequentially one by one and for each component separately in the least square sense. Therefore, for $i,j=1,\dots N$, the objective functions are given by

$$F^{(0)}=\sum_{\mathbf{x}}\left|m^{*(0)}(\mathbf{x})-\rho_Y(\mathbf{x})*m^{(0)}(\mathbf{x})\right|^2, \tag{27a}$$

$$F_i^{(1)}=\sum_{\mathbf{x}}\left|m_i^{*(1)}(\mathbf{x},\tau)-\rho_Y(\mathbf{x})*m^{(1)}(\mathbf{x},\tau)-\sum_{l=1}^{N}Q_{il}(\tau)\frac{\partial}{\partial x_l}m^{*(0)}(\mathbf{x})\right|^2, \tag{27b}$$

$$F_{ij}^{(2)}=\sum_{\mathbf{x}}\left|m_{ij}^{*(2)}(\mathbf{x},\tau)-\rho_Y(\mathbf{x})*m_{ij}^{(2)}(\mathbf{x},\tau)-(Q_{ij}(\tau)+Q_{ji}(\tau))m^{*(0)}(\mathbf{x})-\sum_{l=1}^{N}Q_{il}(\tau)\frac{\partial}{\partial x_l}m_j^{*(1)}(\mathbf{x},\tau)-\sum_{q=1}^{N}Q_{jq}(\tau)\frac{\partial}{\partial x_q}m_i^{*(1)}(\mathbf{x},\tau)+\sum_{l=1}^{N}\sum_{q=1}^{N}Q_{il}(\tau)Q_{jq}(\tau)\frac{\partial}{\partial x_l}\frac{\partial}{\partial x_q}m^{*(0)}(\mathbf{x})\right|^2. \tag{27c}$$

5.2. Regularization

Equations (5) are multidimensional integral equations of the convolution type, which are known to be ill-posed [20] and to deal with this, the mathematical technique of regularization has been employed.

The idea behind it is to introduce additional information in order to solve the problem, here, a restriction for smoothness of the moments was chosen to compensate numerical fluctuations that occur due to binning.

Therefore, a weighted penalty term given by

$$p^{(0)}=\sum_{\mathbf{x}}\sum_{l=1}^{N}\left(\frac{\partial m^{(0)}(\mathbf{x})}{\partial x_l}\right)^2, \tag{28a}$$

$$p_i^{(1)}(\tau)=\sum_{\mathbf{x}}\sum_{l=1}^{N}\left(\frac{\partial m_i^{(1)}(\mathbf{x},\tau)}{\partial x_l}\right)^2, \tag{28b}$$

$$p_{ij}^{(2)}(\tau)=\sum_{\mathbf{x}}\sum_{l=1}^{N}\left(\frac{\partial m_{ij}^{(2)}(\mathbf{x},\tau)}{\partial x_l}\right)^2, \tag{28c}$$

is added to the corresponding objective function.

The weight $\alpha^{(0)},\alpha_i^{(1)},\alpha_{ij}^{(2)}$ with $i,j=1,\dots,N$, strongly influences the outcome of the optimization problem and therefore it is crucial to select an appropriate weight. The strategy employed in this paper is described in the following.

The zeroth joint moment $m^{(0)}(\mathbf{x})$ is known to be constant in τ and all components of the first and second joint moments, $m_i^{(1)}(\mathbf{x},\tau)$ and $m_{ij}^{(2)}(\mathbf{x},\tau)$ respectively, are known to be linear in τ [7,5]. Approximations of each of the joint moments $m'^{(0)}(\mathbf{x},\alpha^{(0)})$, $m_i'^{(1)}(x,\tau,\alpha_i^{(1)})$ and $m_{ij}'^{(2)}(x,\tau,\alpha_{ij}^{(2)})$ are computed for several time increments $\tau=k\Delta t$ with $k\in\mathcal{K}$ and several weights $\alpha^{(0)},\alpha_i^{(1)},\alpha_{ij}^{(2)}\in\mathcal{A}=\{\alpha_{min},\dots,\alpha_{max}\}$. For each penalty weight a linear fit is performed on the numerical solutions in τ, for each $\mathbf{x}$ yielding a straight line $g(\mathbf{x},\alpha,\tau)=a(\mathbf{x},\alpha)+b(\mathbf{x},\alpha)\tau$, for $\alpha\in\{\alpha^{(0)},\alpha_i^{(1)},\alpha_{ij}^{(2)}\}$.

Since the zeroth joint moment $m^{(0)}(\mathbf{x})$ is constant, the corresponding slopes $b(\mathbf{x},\alpha^{(0)})$ should vanish for the appropriate penalty weight. Therefore, to choose the weight for the approximation of the zeroth moment the sum of the absolute values of the slopes over all $\mathbf{x}$ is computed as

$$s^{(0)}(\alpha^{(0)})=\sum_{\mathbf{x}}\left|b(\mathbf{x},\alpha^{(0)})\right|. \tag{29}$$

For the first and second moment that are linear in τ, the squared residuals are computed, weighted with the absolute value of the corresponding slope and summed up over all $\mathbf{x}$:

$$r_i^{(1)}(\mathbf{x},\alpha_i^{(1)},\tau)=\sum_{\mathbf{x}}\left|b(\mathbf{x},\alpha_i^{(1)})\right|\left(g(\mathbf{x},\alpha_i^{(1)},\tau)-m_i'^{(1)}(\mathbf{x},\alpha_i^{(1)},\tau)\right)^2, \tag{30a}$$

$$r_{ij}^{(2)}(\mathbf{x},\alpha_{ij}^{(2)},\tau)=\sum_{\mathbf{x}}\left|b(\mathbf{x},\alpha_{ij}^{(2)})\right|\left(g(\mathbf{x},\alpha_{ij}^{(2)},\tau)-m_{ij}'^{(2)}(\mathbf{x},\alpha_{ij}^{(2)},\tau)\right)^2, \tag{30b}$$

where $i,j=1,\dots,N$.

Choosing the appropriate weight for the penalty term is a trade-off between minimizing the summed slopes (zeroth joint moment) and the weighted residuals (first and second joint moment) of the linear fit and minimizing the difference of left and right hand side of Eqs. (5), which increases after a threshold weight for each of the joint moments: Since an increasing penalty weight leads to flatter moments, both the summed slopes and the summed residuals decrease with increasing weight. If the penalty weight is too large, the penalty term outweighs the corresponding objective function term $F^{(0)},F_i^{(1)}$ and $F_{ij}^{(2)}$. As a consequence, the resulting approximated joint moment does not resemble the empirical joint moment. Therefore, to pick an appropriate weight for the estimation of $m^{(0)}(\mathbf{x})$, the sum of $F^{(0)}$ and $s^{(0)}(\alpha^{(0)})$ is minimized. Typically, for the first and second joint moments, the sum of the weighted residuals $r_i^{(1)}(\mathbf{x},\alpha_i^{(1)},\tau)$, $r_{ij}^{(2)}(\mathbf{x},\alpha_{ij}^{(2)},\tau)$, and the objective function terms $F_i^{(1)}$, $F_{ij}^{(2)}$ are not of the same order of magnitude. Therefore, they are normalized by their maximum value and the offset in $F_i^{(1)}$, $F_{ij}^{(2)}$ is removed. The appropriate weight is given by the one corresponding to the minimum of the sum of the objective function with the residual.

5.3. The final optimization problems

Since $m^{(0)}(\mathbf{x})$ is a probability density function (defined in analogy to its noisy counterpart (see Eq. (5a)), its integral over the full range of $\mathbf{x}$-values equals one. Therefore, the constraint

Please cite this article in press as: T. Scholz et al., Parameter-free resolution of the superposition of stochastic signals, Phys. Lett. A (2016), http://dx.doi.org/10.1016/j.physleta.2016.09.057

$$\int_{\mathbf{x}} m^{(0)}(\mathbf{x})dx_1 \dots dx_N = 1 \tag{31}$$

is added to the optimization problem for the estimation of $m^{(0)}(\mathbf{x})$. The full optimization problem is therefore given by the minimization of $F^{(0)} + \alpha^{(0)} p^{(0)}$, subject to Eq. (31) with positive real $\alpha^{(0)}$.

For the first and second moments no additional constraints are imposed and the final optimization problems are simply given by the minimization of $F_i^{(1)} + \alpha_i^{(1)} p_i^{(1)}$ and $F_{ij}^{(2)} + \alpha_{ij}^{(2)} p_{ij}^{(2)}$ with positive real $\alpha_i^{(1)}, \alpha_{ij}^{(2)}$ and $i, j = 1, \dots, N$.

6. An illustrative example

The previous two sections describe how to obtain the measurement noise parameters **A** and **B** as well as the joint moments $m^0(\mathbf{x})$, $\mathbf{m}^1(\mathbf{x}, \tau)$ and $\mathbf{m}^2(\mathbf{x}, \tau)$. To demonstrate that this framework yields correct results it is tested on two-dimensional synthetic data, which is described in this section. Moreover, the results of solving Eqs. (13) and (5) for this example are presented and discussed.

6.1. The data

Two time-series, each comprising 10^6 data points are generated: $\mathbf{X}(t)$, the stochastic process and $\mathbf{Y}(t)$, the measurement noise. Together they yield the noisy stochastic process $\mathbf{X}^*(t)$ (see Eq. (3)). The time-step between consecutive datapoints is $\Delta t = 0.005$ in arbitrary units.

The time-series of the stochastic process $\mathbf{X}(t)$ is obtained by stochastic integration of Eq. (1), where, with $\mathbf{x} = (x_1, x_2)$

$$\mathbf{D}^{(1)}(\mathbf{x}) = \begin{pmatrix} x_1 - x_1 x_2 \\ x_1^2 - x_2 \end{pmatrix}, \tag{32a}$$

$$\mathbf{D}^{(2)}(\mathbf{x}) = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5(1 + x_1^2) \end{pmatrix}. \tag{32b}$$

The time-series of the measurement noise $\mathbf{Y}$ is obtained by stochastic integration of Eq. (4), where

$$\mathbf{A} = \begin{pmatrix} 200 & -\frac{200}{3} \\ 0 & \frac{200}{3} \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 75 & -\frac{425}{12} \\ -\frac{425}{12} & \frac{125}{6} \end{pmatrix}. \tag{33}$$

From **A** and **B**, the values of the correlation matrix **M** and the covariance matrix **V** are computed as

$$\mathbf{M} = \begin{pmatrix} e^{-1} & \frac{e^{-\frac{1}{3}} - e^{-1}}{2} \\ 0 & e^{-\frac{1}{3}} \end{pmatrix}, \qquad \mathbf{V} = \begin{pmatrix} \frac{5}{32} & -\frac{3}{32} \\ -\frac{3}{32} & \frac{5}{32} \end{pmatrix}, \tag{34}$$

see Eqs. (11a) and (11b). To illustrate the time-series, Fig. 1 shows $\mathbf{X}(t)$ (top), $\mathbf{Y}(t)$ (middle) and $\mathbf{X}^*(t)$ (bottom) for 1000 time-steps.

The ratio $\frac{\sigma_{noise}}{\sigma_{process}}$ between the standard deviation of the measurement noise σ_{noise} and the standard deviation of the stochastic process $\sigma_{process}$ is approximately 0.4 for the first and 0.5 for the second component. For the estimation of the noisy moments $m^{*(0)}(\mathbf{x})$, $\mathbf{m}^{*(1)}(\mathbf{x}, \tau)$ and $\mathbf{m}^{*(2)}(\mathbf{x}, \tau)$ time-steps $\tau = k\Delta t$ with $k = 1, 2, 3, 4, 5, 10, 20, 30, 40, 50$ were chosen.

6.2. Measurement noise parameters

Equation system (13) depends on the number of time-steps k included into the system and the degree ν_{max} of the polynomial in **P**. A determined or overdetermined system is obtained when $\nu_{max} \leq k_{max} - 2$. Therefore, to choose an appropriate ν_{max} for a set $\mathcal{K}$ of time-steps k, the system is solved for different values of ν_{max} up to $k_{max} - 2$ and the values corresponding to the minimal error are chosen. To quantify the error, the 2-norm, D_M and D_V respectively, between the eigenvalues of **M** and **V** and their approximations is used. To apply the method described in Sec. 4, two sets of time-steps involving different time-scales were used, $\mathcal{K}_1 = \{1, 2, 3, 4, 5\}$ and $\mathcal{K}_2 = 5\mathcal{K}_1$. The results are presented in Table 1. Table 1 shows, that the best approximation is obtained for $\mathcal{K}_2$ and $\nu_{max} = 2$. For both sets $\mathcal{K}_1$ and $\mathcal{K}_2$ the worst results are obtained for $\nu_{max} = 3$. However, for all other cases the optimization yields good results, showing that the method is robust with regard to the time scale of $\mathcal{K}$.

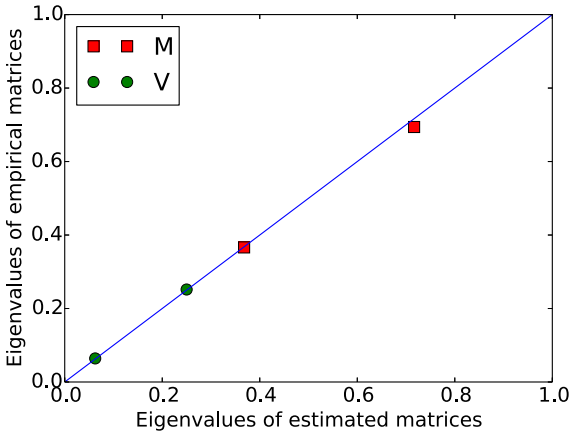


Fig. 3. Eigenvalues of the measurement noise covariance matrix **V** and its matrix of correlation functions **M**, computed analytically and by solving Eq. (13) in the least square sense.

Table 1
Distance D_M and D_V, respectively, between the eigenvalues of **M** and **V** and their approximations for two sets of time-steps $\mathcal{K}_1$ and $\mathcal{K}_2$ and different values of ν_{max}.

	ν_{max}	1	2	3
$\mathcal{K}_1$	D_M	0.024	0.016	0.071
	D_V	0.262	0.265	0.908
$\mathcal{K}_2$	D_M	0.026	0.011	0.168
	D_V	0.001	0.004	0.017

Fig. 3 shows the eigenvalues of the matrices **M** and **V** computed by the optimization (with $\mathcal{K}_2$ and $\nu_{max} = 2$) plotted against their analytical analogues on the left. For comparison, the bisectrix is plotted in blue, indicating perfect agreement between simulation and reconstructed values.

6.3. Joint moments

As an example, this section demonstrates the application of the algorithm to $m^{(0)}$, $m_1^{(1)}$, and $m_{1,1}^{(2)}$.

To approximate the components of the joint moments, the procedure to select an appropriate penalty weight for the optimization layed out in Sec. 5 was carried out for $k \in \{1, 2, 3, 4, 5, 10, 20, 30, 40, 50\}$ and $\alpha^{(0)}, \alpha_i^{(1)}, \alpha_{ij}^{(2)} \in \{0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}$. Fig. 4 shows that a penalty weight of $\alpha^{(0)} = 0.01$ and $\alpha_1^{(1)} = 0.05$ should be selected. For the second moment, the curves show a similar behavior as for the first moment and a weight of $\alpha_{11}^{(2)} = 0.1$ is the appropriate choice (data not shown). To verify that the suggested measures lead to good values of the penalty weights, for each weight the squared sum of the differences between the estimated and the empirical joint moment $MSE^{(0)}$ and $MSE_i^{(1)}$ is computed. Results show, that the

Please cite this article in press as: T. Scholz et al., Parameter-free resolution of the superposition of stochastic signals, Phys. Lett. A (2016), http://dx.doi.org/10.1016/j.physleta.2016.09.057

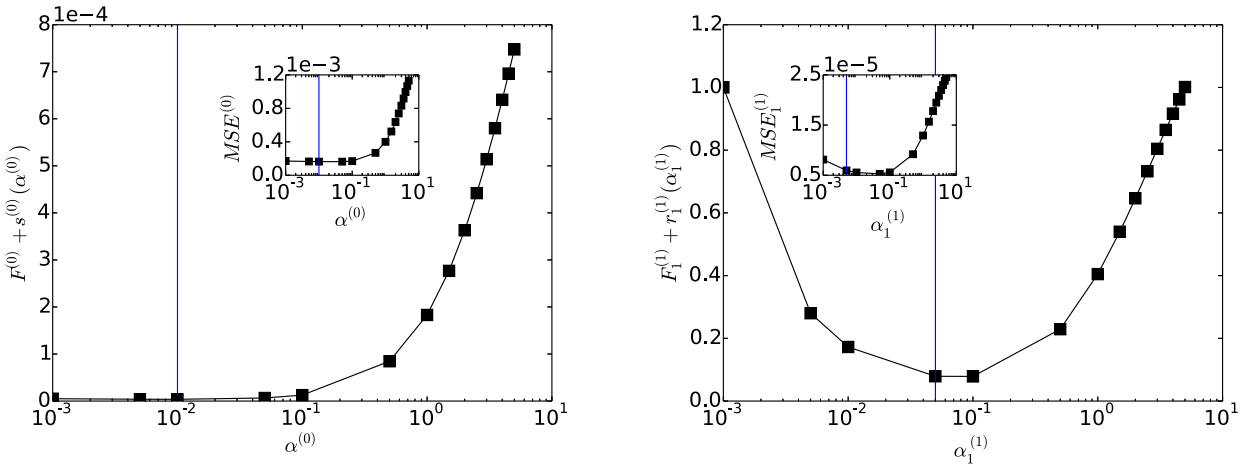


Fig. 4. Illustration of the figures indicating the penalty weight: sum of $F^{(0)}$ and $s^{(0)}(\alpha^{(0)})$ (left) and $r_i^{(1)}(\mathbf{x}, \tau, \alpha_i^{(1)})$ and $F_i^{(1)}$ (right) for the approximations $m'^{(0)}(x, \alpha_0)$ (left) and $m_i'^{(1)}(x, \tau, \alpha_1)$ (right) in dependence of the penalty weight. The inset shows the difference of the estimated and empirical zeroth (left) and first (right) joint moment. For the second moments $m_{ij}'^{(2)}$ results similar to $m_i'^{(1)}$ are obtained.

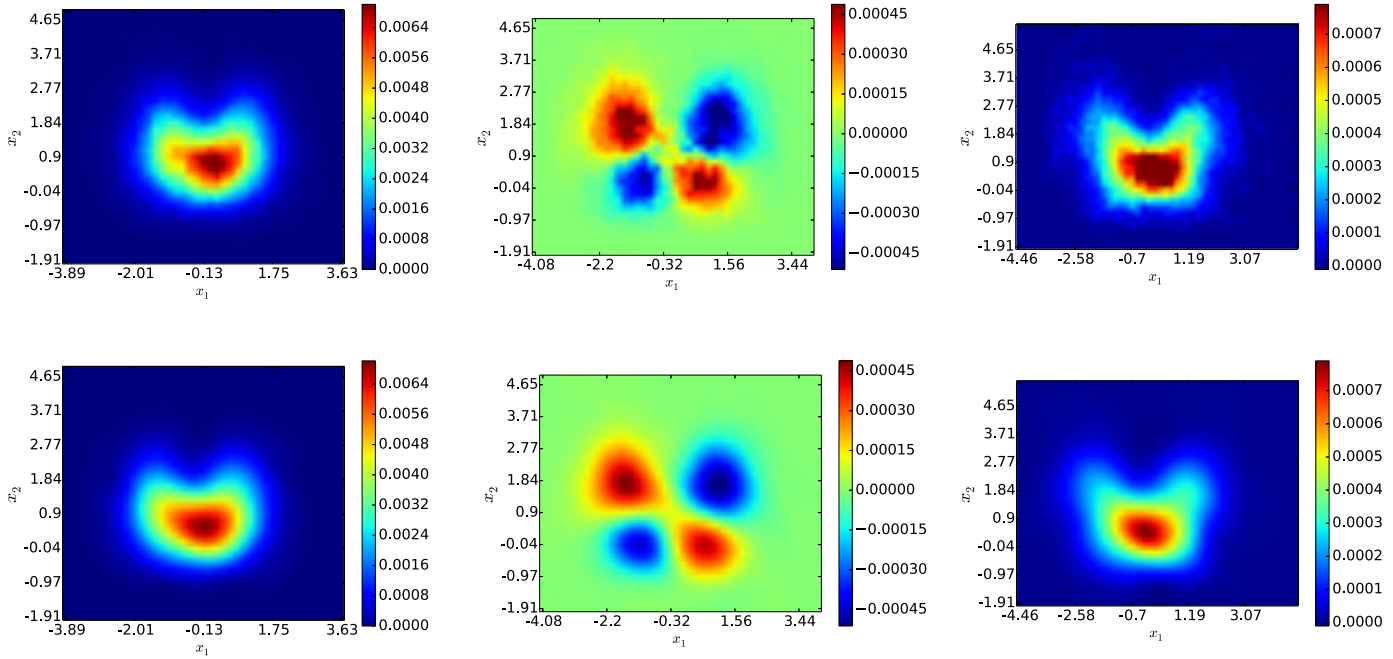


Fig. 5. Zeroth (left), first (middle) and second (right) empirical (top) and estimated joint moments (bottom) with penalty weights 0.01 for the zeroth moment, 0.05 for the first moment and 0.1 for the second joint moment.

penalty weights selected with the method introduced above are a good choice, see insets Fig. 4.

The results of the optimization with the penalty weights determined as described above are shown in Fig. 5. The estimates are smooth and fit well the original data, even in some of its small details.

From the approximated joint moments, using Eqs. (12), the drift and diffusion coefficients were estimated. Fig. 6 presents the estimated (top) and analytical (bottom) first component of the drift coefficient for x_1-values between -5 and 4 and x_2-values between -2 and 5. Furthermore, the first component of the diffusion coefficient whose analytical value is constantly 0.5 in all discretization bins was computed. Good results were obtained, however, as for the drift coefficient, the values on the margins were worse due to unsufficient data in the corresponding bins (data not shown). Thus, the presented method provides satisfactory results for the drift- and diffusion-coefficients of time-series data spoiled with strong measurement noise.

6.4. Limits of the method

The previous section demonstrates that the presented method is able to extract good numerical estimations of the joint moments from a given stochastic process subject to measurement noise, see Fig. 5. To test the limits of this approach, several measurement noise time-series with different statistical properties were created. The two properties tested here are the variation of a) the standard deviation and, b) the time scale of the measurement noise. The proposed method was applied to the generated data and the results are presented and discussed in this section.

The covariance matrix $\mathbf{V}$ of a stationary, two-dimensional Ornstein–Uhlenbeck process $\mathbf{Y}(t)$, described by Eq. (4), is given by

Please cite this article in press as: T. Scholz et al., Parameter-free resolution of the superposition of stochastic signals, Phys. Lett. A (2016), http://dx.doi.org/10.1016/j.physleta.2016.09.057

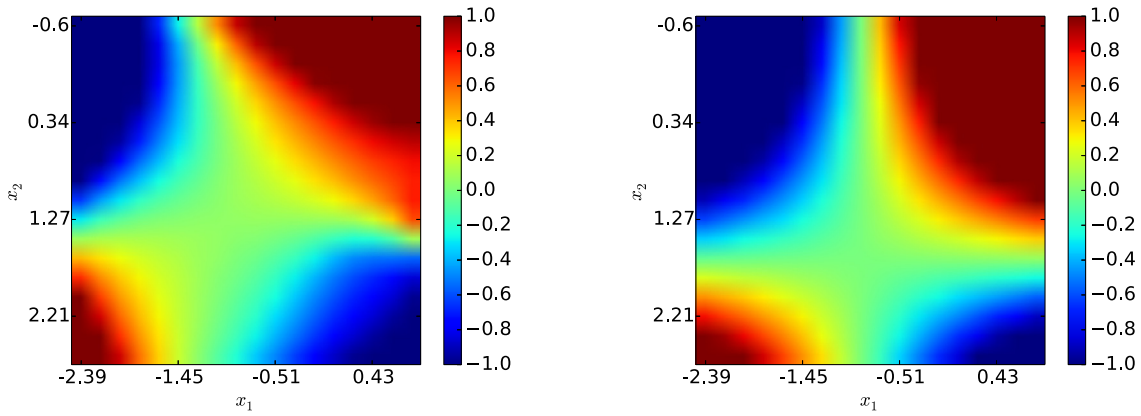


Fig. 6. First component of the approximated (left) and analytical (right) drift coefficient.

Table 2
Distance $MSE^{(0)}(\theta,\gamma)$ of the approximated and the empirical zeroth joint moment for different measurement noise time-series generated with the parameter matrices **A** and **B** scaled by γ and θ, respectively. To vary the standard-deviation, values of $\theta = 0.5, 1, 2$ were chosen while γ was constant 1 (top). To vary the time-scale, both factors were set to 0.5 and 1 (bottom).

θ/γ	0.5	1.0	2.0
$MSE^{(0)}(\theta,1)$	1.55×10^{-4}	1.63×10^{-4}	1.78×10^{-4}
$MSE^{(0)}(\theta,\gamma)$	1.59×10^{-4}	1.63×10^{-4}	–

$$\mathbf{V} = \frac{|\mathbf{A}|\sqrt{\mathbf{B}}\sqrt{\mathbf{B}}^T + \tilde{\mathbf{A}}\sqrt{\mathbf{B}}\sqrt{\mathbf{B}}^T\tilde{\mathbf{A}}^T}{2Tr(\mathbf{A})|\mathbf{A}|}, \tag{35}$$

where $\tilde{\mathbf{A}} = (\mathbf{A} - Tr(\mathbf{A})\mathbf{Id})$ and $Tr(\mathbf{A})$ denotes the trace of **A** [6]. Therefore, to generate time-series that vary in standard-deviation, not in time-scale, it suffices to scale the matrix **B** and keep the matrix **A** unchanged. To vary the time-scale, but not the standard-deviation, it is necessary to scale both matrices **A** and **B** with the same factor.

For this analysis, four additional time-series $\mathbf{Y}_{(\gamma,\theta)}(t)$ were created, where the measurement noise parameters (**A** and **B**) were scaled with factors γ and θ, respectively. The tested scaling factors were $(1, 0.5)$ to investigate the effects of a lower and $(1, 2)$ for a higher standard-deviation than the example presented in Sec. 6. To investigate the limit of the presented method in terms of time-scale, new measurement noise data was generated using scaling factors of $(0.5, 0.5)$ and $(2, 2)$ to yield a slower and a faster process than the Sec. 6 example data, respectively. The generated measurement noise time-series were then added to the stochastic process and the proposed method applied the resulting noisy data.

Table 2 presents the summed square differences between the approximated and empirical zeroth joint moment $MSE^{(0)}$ in terms of variance and time-scale. As expected, the error of approximation increases with increasing standard-deviation of the noise process. However, even for a ratio $\frac{\sigma_{noise}}{\sigma_{process}}$ of 0.75 in the first component and 0.96 in the second, i.e. fluctuations in the measurement noise that are almost as big as the ones in the underlying stochastic process, satisfactory results can be obtained.

In terms of time-scale, a slower measurement noise process yields better results. For the noisy process on a faster time-scale than the example of Sec. 6, no results are presented, since the approximation of the time-correlation matrix **M** through Eqs. (13) was unsuccessful, likely, in this case the fast process could not be captured with the sampling time Δt. A more detailed discussion of the relation between the time-scale of the stochastic processes and the success of the approach is provided by Lehle [8].

7. Conclusions

In conclusion, this paper investigated the possibility to infer the underlying stochastic process and the properties of measurement noise from a N-dimensional measured time-series. The presented approach is based on only three assumptions: a) that the process can be modeled as an Ornstein–Uhlenbeck process, b) that it operates on a faster time-scale than the stochastic process it superimposes and, c) that the two processes are independent.

Moreover, it is parameter-free and thus can be applied to any Markovian multiplicative Gaussian white noise process. Application to synthetic data shows, that the presented method works for a wide range of amplitude ratios of the stochastic process and the measurement noise. In addition, the reconstruction succeeds with high accuracy for different time-scales of the measurement noise process. The algorithm requires solely a standard PC without special software environment and solves the inversion problem within the order of minutes.

In general, our method can be taken for denoising of measurement sets of stochastic variables subjected to strong measurement noise. As discussed in the introduction, the ability was previously introduced in simpler scenarios and here we solved the problem without parameter tunning and assuming measurement noise sources with correlations in time.

The computational implementation of our method can now be applied to sets of empirical data. The are appealing applications with geophysical data, namely using pressure series defining spatial patterns in large-scale vacillation in atmospheric masses [10], e.g. El Niño Southern Oscillation or the North Atlantic Oscillation. Here measurement noise was previously detected [3]. Other important applications intersect the field of medicine and biophysics, namely investigating the coupling of different EEG signals taken from different points on the skalp, which introduces considerable amounts of measurement noise in the data [14]. In a forthcoming investigation, we intend to apply this method to data sets of lift and drag measurements taken from a blade in a wind tunnel, considering different angles of attack [11].

Acknowledgements

The authors thank Fundação para a Ciência e a Tecnologia (FCT) for financial support under UID/FIS/00618/2013, UID/MAT/04561/2013, FCOMP-01-0124-FEDER-016080, SFRH/BPD/65427/2009 (FR) and SFRH/BD/86934/2012 (TS). PGL thanks the German Environment Ministry (0325577B) and the German Foundation for Research (DFG) for financial support (MA 1636/9-1). TS thanks the IPID4all programme supported by the German Academic Exchange Service (DAAD) with funds from the Federal Ministry of Education

and Research (BMBF). This work is part of a bilateral cooperation DRI/DAAD/1208/2013 supported by FCT and DAAD. VVL thanks the Prometeo project of SENESCYT (Ecuador) for financial support.

Appendix A. Relation between the joint moments and their noisy analogues

The calculations leading to Eqs. (5) are given in this appendix as an expanded version of the derivation by [8]. For notational simplicity the following abbreviation for the multivariate normal probability density function is used:

$$\mathcal{G}(\mathbf{x},\mu,\Sigma)=\frac{1}{\sqrt{(2\pi)^N|det(\Sigma)|}}e^{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)} \tag{A.1}$$

where $\mu\in\mathbb{R}^N$ is the mean of the distribution and $\Sigma\in\mathbb{R}^{N\times N}$ is the covariance matrix of the variable $\mathbf{X}$. Thus, the one and two point probability density functions of the measurement noise $\mathbf{Y}$ can be written as

$$\rho_Y(\mathbf{x})=\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{V}) \tag{A.2a}$$

$$\rho_Y(\mathbf{x},\mathbf{x}',\tau)=\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{V})\mathcal{G}(\mathbf{x}',\mathbf{M}(\tau)\mathbf{x},\mathbf{C}), \tag{A.2b}$$

where $\mathbf{C}=\mathbf{V}-\mathbf{M}(\tau)\mathbf{V}(\mathbf{M}(\tau))^T$.

In analogy to Eq. (5) the conditional moments of the measurement noise free process are defined as

$$m^{(0)}(\mathbf{x})=\int_{\mathbf{x}'}\rho(\mathbf{x},\mathbf{x}',\tau)d\mathbf{x}', \tag{A.3a}$$

$$m_i^{(1)}(\mathbf{x},\tau)=\int_{\mathbf{x}'}(x_i'(t+\tau)-x_i(t))\rho(\mathbf{x},\mathbf{x}',\tau)d\mathbf{x}', \tag{A.3b}$$

$$m_{ij}^{(2)}(\mathbf{x},\tau)=\int_{\mathbf{x}'}(x_i'(t+\tau)-x_i(t))(x_j'(t+\tau)-x_j(t))\rho(\mathbf{x},\mathbf{x}',\tau)d\mathbf{x}'. \tag{A.3c}$$

The assumption that the measurement noise and the underlying stochastic process are independent yields $\rho^*(\mathbf{x},\mathbf{x}',\tau)=\rho(\mathbf{x},\mathbf{x}',\tau)*\rho_Y(\mathbf{x},\mathbf{x}',\tau)$. Using this relation, as well as Eqs. (A.2) and the notation $d\mathbf{x}=dx_1...dx_N$, the derivation of Eq. (5a) reads:

$$\begin{aligned}
m^{*(0)}(\mathbf{x})&=\int_{\mathbf{x}'}\rho^*(\mathbf{x},\mathbf{x}',\tau)d\mathbf{x}'\\
&=\int_{\mathbf{x}'}\int_{\mathbf{z}}\int_{\mathbf{z}'}\rho(\mathbf{z},\mathbf{z}',\tau)\rho_Y(\mathbf{x}-\mathbf{z},\mathbf{x}'-\mathbf{z}',\tau)d\mathbf{z}d\mathbf{z}'d\mathbf{x}'\\
&=\int_{\mathbf{x}'}\int_{\mathbf{z}}\int_{\mathbf{z}'}\rho(\mathbf{z},\mathbf{z}',\tau)\mathcal{G}(\mathbf{x}-\mathbf{z},0,\mathbf{V})\mathcal{G}(\mathbf{x}'-\mathbf{z}',\mathbf{M}(\tau)\\
&\quad\times(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{z}d\mathbf{z}'d\mathbf{x}'\\
&=\int_{\mathbf{x}'}\int_{\mathbf{z}}\int_{\mathbf{z}'}\rho(\mathbf{z},\mathbf{z}',\tau)\mathcal{G}(\mathbf{x},\mathbf{z},\mathbf{V})\mathcal{G}(\mathbf{x}',\mathbf{z}'\\
&\quad+\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{z}d\mathbf{z}'d\mathbf{x}',
\end{aligned}$$

where in the last step the variables $\mathbf{x}$ and $\mathbf{z}$ have been shifted by $\mathbf{z}$ and $\mathbf{z}'$, respectively. Thus,

$$\begin{aligned}
m^{*(0)}(\mathbf{x})&=\int_{\mathbf{z}}\mathcal{G}(\mathbf{x},\mathbf{z},\mathbf{V})\int_{\mathbf{z}'}\rho(\mathbf{z},\mathbf{z}',\tau)\\
&\quad\times\int_{\mathbf{x}'}\mathcal{G}(\mathbf{x}',\mathbf{z}'+\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{x}'d\mathbf{z}'d\mathbf{z}\\
&=\int_{\mathbf{z}}\mathcal{G}(\mathbf{x},\mathbf{z},\mathbf{V})\int_{\mathbf{z}'}\rho(\mathbf{z},\mathbf{z}',\tau)d\mathbf{z}'d\mathbf{z}\\
&=\int_{\mathbf{z}}\mathcal{G}(\mathbf{x},\mathbf{z},\mathbf{V})m^{(0)}(\mathbf{z})d\mathbf{z}\\
&=\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{V})*m^{(0)}(\mathbf{x})=\rho_Y(\mathbf{x})*m^{(0)}(\mathbf{x}),
\end{aligned} \tag{A.4}$$

using $\int_{\mathbf{x}'}\mathcal{G}(\mathbf{x}',\mathbf{z}'+\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{x}'=1$ and Eq. (A.3a). With the same abbreviations as above, the derivation of Eq. (5b) is similar:

$$\begin{aligned}
m_i^{(1)}(\mathbf{x},\tau)&=\int_{\mathbf{x}'}(x_i'-x_i)\rho^*(\mathbf{x},\mathbf{x}',\tau)d\mathbf{x}'\\
&=\int_{\mathbf{x}'}(x_i'-x_i)\int_{\mathbf{z}}\int_{\mathbf{z}'}\rho(\mathbf{z},\mathbf{z}',\tau)\rho_Y(\mathbf{x}-\mathbf{z},\mathbf{x}'-\mathbf{z}',\tau)d\mathbf{z}d\mathbf{z}'d\mathbf{x}'\\
&=\int_{\mathbf{x}'}(x_i'-x_i)\int_{\mathbf{z}}\int_{\mathbf{z}'}\rho(\mathbf{z},\mathbf{z}',\tau)\mathcal{G}(\mathbf{x},\mathbf{z},\mathbf{V})\mathcal{G}(\mathbf{x}',\mathbf{z}'\\
&\quad+\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{z}d\mathbf{z}'d\mathbf{x}'\\
&=\int_{\mathbf{z}}\mathcal{G}(\mathbf{x},\mathbf{z},\mathbf{V})\int_{\mathbf{z}'}\rho(\mathbf{z},\mathbf{z}',\tau)\int_{\mathbf{x}'}(x_i'-x_i)\mathcal{G}(\mathbf{x}',\mathbf{z}'\\
&\quad+\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{x}'d\mathbf{z}'d\mathbf{z}\\
&=\int_{\mathbf{z}}\mathcal{G}(\mathbf{x},\mathbf{z},\mathbf{V})\int_{\mathbf{z}'}\rho(\mathbf{z},\mathbf{z}',\tau)\left(\int_{\mathbf{x}'}x_i'\mathcal{G}(\mathbf{x}',\mathbf{z}'+\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{x}'\right.\\
&\quad\left.-x_i\int_{\mathbf{x}'}\mathcal{G}(\mathbf{x}',\mathbf{z}'+\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{x}'\right)d\mathbf{z}'d\mathbf{z}\\
&=\int_{\mathbf{z}}\mathcal{G}(\mathbf{x},\mathbf{z},\mathbf{V})\int_{\mathbf{z}'}\rho(\mathbf{z},\mathbf{z}',\tau)\left(\int_{\mathbf{x}'}x_i'\mathcal{G}(\mathbf{x}',\mathbf{z}'\right.\\
&\quad\left.+\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{x}'-x_i\right)d\mathbf{z}'d\mathbf{z},
\end{aligned}$$

where again a variable shift and $\int_{\mathbf{x}'}\mathcal{G}(\mathbf{x}',\mathbf{z}'+\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{x}'=1$ was used. Noting that $\int_{\mathbf{x}'}x_i'\mathcal{G}(\mathbf{x}',\mathbf{z}'+\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{x}'$ is the i-th component of the mean value of the Gaussian function it follows that

$$\begin{aligned}
m_i^{(1)}(\mathbf{x},\tau)&=\int_{\mathbf{z}}\mathcal{G}(\mathbf{x},\mathbf{z},\mathbf{V})\int_{\mathbf{z}'}\rho(\mathbf{z},\mathbf{z}',\tau)\\
&\quad\times\left(z_i'+\sum_{l=1}^{N}M_{il}(\tau)(x_l-z_l)-x_i\right)d\mathbf{z}'d\mathbf{z}\\
&=\int_{\mathbf{z}}\mathcal{G}(\mathbf{x},\mathbf{z},\mathbf{V})\int_{\mathbf{z}'}\rho(\mathbf{z},\mathbf{z}',\tau)\left(z_i'-z_i+z_i-x_i\right.\\
&\quad\left.+\sum_{l=1}^{N}M_{il}(\tau)(x_l-z_l)\right)d\mathbf{z}'d\mathbf{z}\\
&=\int_{\mathbf{z}}\mathcal{G}(\mathbf{x},\mathbf{z},\mathbf{V})\left(\int_{\mathbf{z}'}(z_i'-z_i)\rho(\mathbf{z},\mathbf{z}',\tau)d\mathbf{z}'\right.
\end{aligned}$$

Please cite this article in press as: T. Scholz et al., Parameter-free resolution of the superposition of stochastic signals, Phys. Lett. A (2016), http://dx.doi.org/10.1016/j.physleta.2016.09.057

$$
\begin{aligned}
&+\int_{\mathbf{z}'}\left(z_i - x_i + \sum_{l=1}^{N} M_{il}(\tau)(x_l - z_l)\right)\rho(\mathbf{z},\mathbf{z}',\tau)d\mathbf{z}'\Bigg)d\mathbf{z}\\
&=\int_{\mathbf{z}}\mathcal{G}(\mathbf{x},\mathbf{z},\mathbf{V})\Bigg(m_i^{(1)}(\mathbf{z},\tau)+\Bigg(z_i - x_i + M_{ii}(\tau)(x_i - z_i)\\
&\quad+\sum_{l\neq i} M_{il}(\tau)(x_l - z_l)\Bigg)\int_{\mathbf{z}'}\rho(\mathbf{z},\mathbf{z}',\tau)d\mathbf{z}'\Bigg)d\mathbf{z}\\
&=\int_{\mathbf{z}}\mathcal{G}(\mathbf{x},\mathbf{z},\mathbf{V})\Bigg(m_i^{(1)}(\mathbf{z},\tau)\\
&\quad-\sum_{l=1}^{N}(\delta_{il}-M_{il})(x_l - z_l)m^{(0)}(\mathbf{z})\Bigg)d\mathbf{z}\\
&=\int_{\mathbf{z}}\mathcal{G}(\mathbf{x},\mathbf{z},\mathbf{V})m_i^{(1)}(\mathbf{z},\tau)d\mathbf{z}\\
&\quad-\int_{\mathbf{z}}\mathcal{G}(\mathbf{x},\mathbf{z},\mathbf{V})\sum_{l=1}^{N}(\delta_{il}-M_{il})(x_l - z_l)m^{(0)}(\mathbf{z})d\mathbf{z}\\
&=\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{V})*m_i^{(1)}(\mathbf{x},\tau)\\
&\quad-\int_{\mathbf{z}}\sum_{l=1}^{N}(\delta_{il}-M_{il})(x_l - z_l)\mathcal{G}(\mathbf{x},\mathbf{z},\mathbf{V})m^{(0)}(\mathbf{z})d\mathbf{z},
\end{aligned}
\tag{A.5}
$$

where the definitions of the conditional moments (see Eq. (A.3a) and (A.3b)) were used. Applying the definition of the convolution, the integral in the last equation reads

$$
\begin{aligned}
&\int_{\mathbf{z}}\mathcal{G}(\mathbf{x},\mathbf{z},\mathbf{V})\sum_{l=1}^{N}(\delta_{il}-M_{il})(x_l - z_l)m^{(0)}(\mathbf{z})d\mathbf{z}\\
&=\sum_{l=1}^{N}(\delta_{il}-M_{il})\left(x_l\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{V})\right)*m^{(0)}(\mathbf{x})\\
&=-\sum_{l=1}^{N}(\delta_{il}-M_{il})\left(\sum_{j=1}^{N}-\delta_{lj}x_j\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{V})\right)*m^{(0)}(\mathbf{x})\\
&=-\sum_{l=1}^{N}(\delta_{il}-M_{il})\left(\sum_{k=1}^{N}V_{lk}\left(\sum_{j=1}^{N}-\mathbf{V}_{kj}^{-1}x_j\right)\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{V})\right)\\
&\quad*m^{(0)}(\mathbf{x})\\
&=-\sum_{l=1}^{N}(\delta_{il}-M_{il})\left(\sum_{k=1}^{N}V_{lk}\left(\frac{\partial}{\partial x_k}\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{V})\right)\right)\\
&\quad*m^{(0)}(\mathbf{x})\\
&=-\sum_{l=1}^{N}(\delta_{il}-M_{il})\left(\sum_{k=1}^{N}V_{lk}\frac{\partial}{\partial x_k}\left(\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{V})*m^{(0)}(\mathbf{x})\right)\right)\\
&=-\sum_{k=1}^{N}\mathbf{Q}_{ik}\frac{\partial}{\partial x_k}m^{*(0)}(\mathbf{x})
\end{aligned}
\tag{A.6}
$$

where the relation $\frac{\partial}{\partial x_k}\mathcal{G}(\mathbf{x},\mu,\Sigma)=\sum_{l=1}^{N}-\Sigma_{\mathbf{l}}^{-\mathbf{1}}x_l\mathcal{G}(\mathbf{x},\mu,\Sigma)$ was used together with the property of the convolution operation $\frac{\partial}{\partial x_k}\left(\mathcal{G}(\mathbf{x},\mu,\Sigma)*m^{(0)}(\mathbf{x})\right)=\left(\frac{\partial}{\partial x_k}\mathcal{G}(\mathbf{x},\mu,\Sigma)\right)*m^{(0)}(\mathbf{x})$. Introducing Eq. (A.6) in Eq. (A.5) yields the last term in Eq. (5b) as we aim to demonstrate.

Equation (5c) is obtained as follows:

$$
\begin{aligned}
&m_{ij}^{(2)}(\mathbf{x},\tau)\\
&=\int_{\mathbf{x}'}(x_i' - x_i)(x_j' - x_j)\rho^*(\mathbf{x},\mathbf{x}',\tau)d\mathbf{x}'\\
&=\int_{\mathbf{x}'}(x_i' - x_i)(x_j' - x_j)\\
&\quad\times\int_{\mathbf{z}}\int_{\mathbf{z}'}\rho(\mathbf{z},\mathbf{z}',\tau)\rho_Y(\mathbf{x}-\mathbf{z},\mathbf{x}'-\mathbf{z}',\tau)d\mathbf{z}d\mathbf{z}'d\mathbf{x}'\\
&=\int_{\mathbf{x}'}(x_i' - x_i)(x_j' - x_j)\int_{\mathbf{z}}\int_{\mathbf{z}'}\rho(\mathbf{z},\mathbf{z}',\tau)\mathcal{G}(\mathbf{x},\mathbf{z},\mathbf{V})\mathcal{G}(\mathbf{x}',\mathbf{z}'\\
&\quad+\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{z}d\mathbf{z}'d\mathbf{x}'\\
&=\int_{\mathbf{z}}\mathcal{G}(\mathbf{x},\mathbf{z},\mathbf{V})\int_{\mathbf{z}'}\rho(\mathbf{z},\mathbf{z}',\tau)\int_{\mathbf{x}'}(x_i' - x_i)(x_j' - x_j)\mathcal{G}(\mathbf{x}',\mathbf{z}'\\
&\quad+\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{x}'d\mathbf{z}'d\mathbf{z}.
\end{aligned}
\tag{A.7}
$$

The integral over $\mathbf{x}'$ reads

$$
\begin{aligned}
&\int_{\mathbf{x}'}(x_i' - x_i)(x_j' - x_j)\mathcal{G}(\mathbf{x}',\mathbf{z}'+\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{x}'\\
&=\int_{\mathbf{x}'}x_i'x_j'\mathcal{G}(\mathbf{x}',\mathbf{z}'+\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{x}'\\
&\quad-x_j\int_{\mathbf{x}'}x_i'\mathcal{G}(\mathbf{x}',\mathbf{z}'+\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{x}'\\
&\quad-x_i\int_{\mathbf{x}'}x_j'\mathcal{G}(\mathbf{x}',\mathbf{z}'+\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{x}'\\
&\quad+x_ix_j\int_{\mathbf{x}'}\mathcal{G}(\mathbf{x}',\mathbf{z}'+\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{x}'\\
&=\int_{\mathbf{x}'}x_i'x_j'\mathcal{G}(\mathbf{x}',\mathbf{z}'+\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{x}'\\
&\quad-x_j\left(z_i'+\sum_{l=1}^{N}M_{il}(x_l - z_l)\right)\\
&\quad-x_i\left(z_j'+\sum_{q=1}^{N}M_{jq}(x_q - z_q)\right)+x_ix_j,
\end{aligned}
\tag{A.8}
$$

where once again the relations $\int_{\mathbf{x}'}\mathcal{G}(\mathbf{x}',\mathbf{z}'+\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{x}'=1$ and $\int_{\mathbf{x}'}x_i'\mathcal{G}(\mathbf{x}',\mathbf{z}'+\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{x}'=z_i'+\sum_{l=1}^{N}M_{il}(x_l - z_l)$ were used. Using the variable transformation $\mathbf{u}=\mathbf{x}'-\mathbf{z}'-\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z})$, the remaining integral over $\mathbf{x}'$ can be written as

$$
\begin{aligned}
&\int_{\mathbf{x}'}x_i'x_j'\mathcal{G}(\mathbf{x}',\mathbf{z}'+\mathbf{M}(\tau)(\mathbf{x}-\mathbf{z}),\mathbf{C})d\mathbf{x}'\\
&=\int_{\mathbf{u}}\left(u_i+z_i'+\sum_{l=1}^{N}M_{il}(x_l - z_l)\right)\\
&\quad\times\left(u_j+z_j'+\sum_{q=1}^{N}M_{jq}(x_q - z_q)\right)\mathcal{G}(\mathbf{u},\mathbf{0},\mathbf{C})d\mathbf{u}
\end{aligned}
$$

Please cite this article in press as: T. Scholz et al., Parameter-free resolution of the superposition of stochastic signals, Phys. Lett. A (2016), http://dx.doi.org/10.1016/j.physleta.2016.09.057

$$
\begin{aligned}
&= \int_{\mathbf{u}} u_i u_j \mathcal{G}(\mathbf{u}, \mathbf{0}, \mathbf{C}) d\mathbf{u} + \left(z'_i + \sum_{l=1}^{N} M_{il}(x_l - z_l) \right) \\
&\quad \times \int_{\mathbf{u}} u_j \mathcal{G}(\mathbf{u}, \mathbf{0}, \mathbf{C}) d\mathbf{u} \\
&\quad + \left(z'_j + \sum_{q=1}^{N} M_{jq}(x_q - z_q) \right) \int_{\mathbf{u}} u_i \mathcal{G}(\mathbf{u}, \mathbf{0}, \mathbf{C}) d\mathbf{u} \\
&\quad + \left(z'_i + \sum_{l=1}^{N} M_{il}(x_l - z_l) \right) \left(z'_j + \sum_{q=1}^{N} M_{jq}(x_q - z_q) \right) \\
&\quad \times \int_{\mathbf{u}} \mathcal{G}(\mathbf{u}, \mathbf{0}, \mathbf{C}) d\mathbf{u} \\
&= C_{ij} + \left(z'_i + \sum_{l=1}^{N} M_{il}(x_l - z_l) \right) \left(z'_j + \sum_{q=1}^{N} M_{jq}(x_q - z_q) \right),
\end{aligned}
$$

where the first two moments of the Gauss function $\int_{\mathbf{x}} x_i x_j \mathcal{G}(\mathbf{x}, \mathbf{0}, \Sigma) d\mathbf{x} = \Sigma_{ij}$ and $\int_{\mathbf{x}} x_i \mathcal{G}(\mathbf{x}, \mathbf{0}, \Sigma) d\mathbf{x} = 0$ were substituted. Thus, Eq. (A.8) can now be formulated as

$$
\begin{aligned}
&\int_{\mathbf{x}'} (x'_i - x_i)(x'_j - x_j) \mathcal{G}(\mathbf{x}', \mathbf{z}' + \mathbf{M}(\tau)(\mathbf{x} - \mathbf{z}), \mathbf{C}) d\mathbf{x}' \\
&= C_{ij} + \left(z'_i - x_i + \sum_{l=1}^{N} M_{il}(x_l - z_l) \right) \\
&\quad \times \left(z'_j - x_j + \sum_{q=1}^{N} M_{jq}(x_q - z_q) \right),
\end{aligned}
$$

which for the equation of the second moment yields

$$
\begin{aligned}
&m^{(2)}_{ij}(\mathbf{x}, \tau) \\
&= \int_{\mathbf{z}} \mathcal{G}(\mathbf{x}, \mathbf{z}, \mathbf{V}) \int_{\mathbf{z}'} \rho(\mathbf{z}, \mathbf{z}', \tau) \Bigg[C_{ij} \\
&\quad + \left(z'_i - z_i + z_i - x_i + \sum_{l=1}^{N} M_{il}(x_l - z_l) \right) \\
&\quad \times \left(z'_j - z_j + z_j - x_j + \sum_{q=1}^{N} M_{jq}(x_q - z_q) \right) \Bigg] d\mathbf{z}' d\mathbf{z} \\
&= \int_{\mathbf{z}} \mathcal{G}(\mathbf{x}, \mathbf{z}, \mathbf{V}) \Bigg[C_{ij} \int_{\mathbf{z}'} \rho(\mathbf{z}, \mathbf{z}', \tau) d\mathbf{z}' \\
&\quad + \int_{\mathbf{z}'} (z'_i - z_i)(z'_j - z_j) \rho(\mathbf{z}, \mathbf{z}', \tau) d\mathbf{z}' \\
&\quad + \int_{\mathbf{z}'} (z'_i - z_i) \left(z_j - x_j + \sum_{q=1}^{N} M_{jq}(x_q - z_q) \right) \rho(\mathbf{z}, \mathbf{z}', \tau) d\mathbf{z}' \\
&\quad + \int_{\mathbf{z}'} (z'_j - z_j) \left(z_i - x_i + \sum_{l=1}^{N} M_{il}(x_l - z_l) \right) \rho(\mathbf{z}, \mathbf{z}', \tau) d\mathbf{z}' \\
&\quad + \int_{\mathbf{z}'} \left(z_j - x_j + \sum_{q=1}^{N} M_{jq}(x_q - z_q) \right) \\
&\quad \times \left(z_i - x_i + \sum_{l=1}^{N} M_{il}(x_l - z_l) \right) \rho(\mathbf{z}, \mathbf{z}', \tau) d\mathbf{z}' \Bigg] d\mathbf{z} \\
&= \int_{\mathbf{z}} \mathcal{G}(\mathbf{x}, \mathbf{z}, \mathbf{V}) \Bigg[C_{ij} m^{(0)}(\mathbf{z}) + m^{(2)}_{ij}(\mathbf{z}, \tau) + m^{(1)}_i(\mathbf{z}, \tau) \\
&\quad \times \sum_{q=1}^{N} (-\delta_{jq} + M_{jq})(x_q - z_q) \\
&\quad + m^{(1)}_j(\mathbf{z}, \tau) \sum_{l=1}^{N} (-\delta_{il} + M_{il})(x_l - z_l) \\
&\quad + m^{(0)}(\mathbf{z}) \left(\sum_{l=1}^{N} (-\delta_{il} + M_{il})(x_l - z_l) \right) \\
&\quad \times \left(\sum_{q=1}^{N} (-\delta_{jq} + M_{jq})(x_q - z_q) \right) \Bigg] d\mathbf{z},
\end{aligned}
$$

where the definition of the conditional moments (see Eqs. (A.3a), (A.3b) and (A.3c)) were inserted. The definition of the convolution and the relation $\int_{\mathbf{z}} (x_i - z_i) f(\mathbf{x} - \mathbf{z}) g(\mathbf{z}) d\mathbf{z} = (x_i f(\mathbf{x})) * g(\mathbf{x})$ yield

$$
\begin{aligned}
&m^{(2)}_{ij}(\mathbf{x}, \tau) \\
&= C_{ij} \mathcal{G}(\mathbf{x}, \mathbf{0}, \mathbf{V}) * m^{(0)}(\mathbf{x}) + \mathcal{G}(\mathbf{x}, \mathbf{0}, \mathbf{V}) * m^{(2)}_{ij}(\mathbf{x}, \tau) \\
&\quad + \left(\sum_{q=1}^{N} (-\delta_{jq} + M_{jq}) \Big(x_q \mathcal{G}(\mathbf{x}, \mathbf{0}, \mathbf{V}) \Big) \right) * m^{(1)}_i(\mathbf{x}, \tau) \\
&\quad + \left(\sum_{l=1}^{N} (-\delta_{il} + M_{il}) \Big(x_l \mathcal{G}(\mathbf{x}, \mathbf{0}, \mathbf{V}) \Big) \right) * m^{(1)}_j(\mathbf{x}, \tau) \\
&\quad + \left(\sum_{l=1}^{N} (-\delta_{il} + M_{il}) \sum_{q=1}^{N} (-\delta_{jq} + M_{jq}) \Big(x_l x_q \mathcal{G}(\mathbf{x}, \mathbf{0}, \mathbf{V}) \Big) \right) \\
&\quad * m^{(0)}(\mathbf{x}) \\
&= \mathcal{G}(\mathbf{x}, \mathbf{0}, \mathbf{V}) * m^{(2)}_{ij}(\mathbf{x}, \tau) \\
&\quad + \left(\sum_{q=1}^{N} \sum_{k=1}^{N} (\delta_{jq} - M_{jq}) V_{qk} \left(\frac{\partial}{\partial x_k} \mathcal{G}(\mathbf{x}, \mathbf{0}, \mathbf{V}) \right) \right) * m^{(1)}_i(\mathbf{x}, \tau) \\
&\quad + \left(\sum_{l=1}^{N} \sum_{k=1}^{N} (\delta_{il} - M_{il}) V_{lk} \left(\frac{\partial}{\partial x_k} \mathcal{G}(\mathbf{x}, \mathbf{0}, \mathbf{V}) \right) \right) * m^{(1)}_j(\mathbf{x}, \tau) \\
&\quad + \left(C_{ij} + \sum_{l=1}^{N} \sum_{q=1}^{N} (-\delta_{il} + M_{il})(-\delta_{jq} + M_{jq}) V_{lq} \right) \mathcal{G}(\mathbf{x}, \mathbf{0}, \mathbf{V}) \\
&\quad * m^{(0)}(\mathbf{x}) \\
&\quad + \left(\sum_{l=1}^{N} \sum_{q=1}^{N} \sum_{k=1}^{N} \sum_{m=1}^{N} (\delta_{il} - M_{il}) V_{lk} (\delta_{jq} - M_{jq}) V_{qm} \right. \\
&\quad \left. \times \left(\frac{\partial}{\partial x_k} \frac{\partial}{\partial x_m} \mathcal{G}(\mathbf{x}, \mathbf{0}, \mathbf{V}) \right) \right) * m^{(0)}(\mathbf{x}) \\
&= \mathcal{G}(\mathbf{x}, \mathbf{0}, \mathbf{V}) * m^{(2)}_{ij}(\mathbf{x}, \tau) + \left(\sum_{q=1}^{N} (\delta_{jq} - M_{jq}) V_{iq} \right. \\
&\quad \left. + \sum_{l=1}^{N} (\delta_{il} - M_{il}) V_{lj} \right) \mathcal{G}(\mathbf{x}, \mathbf{0}, \mathbf{V}) * m^{(0)}(\mathbf{x})
\end{aligned}
$$

Please cite this article in press as: T. Scholz et al., Parameter-free resolution of the superposition of stochastic signals, Phys. Lett. A (2016), http://dx.doi.org/10.1016/j.physleta.2016.09.057

$$+\left(\sum_{q=1}^{N}\sum_{k=1}^{N}(\delta_{jq}-M_{jq})V_{qk}\left(\frac{\partial}{\partial x_k}\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{V})\right)\right)*m_i^{(1)}(\mathbf{x},\tau)$$
$$+\left(\sum_{l=1}^{N}\sum_{k=1}^{N}(\delta_{il}-M_{il})V_{lk}\left(\frac{\partial}{\partial x_k}\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{V})\right)\right)*m_j^{(1)}(\mathbf{x},\tau)$$
$$+\left(\sum_{l=1}^{N}\sum_{q=1}^{N}\sum_{k=1}^{N}\sum_{m=1}^{N}(\delta_{il}-M_{il})V_{lk}(\delta_{jq}-M_{jq})V_{qm}\times\left(\frac{\partial}{\partial x_k}\frac{\partial}{\partial x_m}\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{V})\right)\right)*m^{(0)}(\mathbf{x}), \tag{A.9}$$

where the derivations of the Gauss functions $x_l\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{\Sigma}) = -\sum_{k=1}^{N}\Sigma_{lk}\frac{\partial}{\partial x_k}\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{\Sigma})$ and $x_lx_q\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{\Sigma}) = \sum_{k=1}^{N}\sum_{m=1}^{N}\Sigma_{lk}\Sigma_{qm}\frac{\partial^2}{\partial x_k\partial x_m}\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{\Sigma}) + \Sigma_{lq}\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{\Sigma})$ were used as well as the following auxiliary calculation

$$\begin{aligned}
&\sum_{l=1}^{N}\sum_{q=1}^{N}(\delta_{il}-M_{il})(\delta_{jq}-M_{jq})V_{lq}+C_{ij}\\
&=\sum_{l=1}^{N}\sum_{q=1}^{N}(\delta_{il}-M_{il})(\delta_{jq}-M_{jq})V_{lq}+V_{ij}\\
&\quad-\sum_{l=1}^{N}M_{il}\left(\sum_{q=1}^{N}V_{lq}M_{qj}^{T}\right)\\
&=\left(V_{iq}-\sum_{l=1}^{N}M_{il}V_{lq}\right)\sum_{q=1}^{N}(\delta_{jq}-M_{jq})+V_{ij}\\
&\quad-\sum_{l=1}^{N}\sum_{q=1}^{N}M_{il}V_{lq}M_{jq}\\
&=V_{ij}-\sum_{l=1}^{N}M_{il}V_{lj}-\sum_{q=1}^{N}V_{iq}M_{jq}+\sum_{l=1}^{N}\sum_{q=1}^{N}M_{il}V_{lq}M_{jq}+V_{ij}\\
&\quad-\sum_{l=1}^{N}\sum_{q=1}^{N}M_{il}V_{lq}M_{jq}\\
&=V_{ij}-\sum_{l=1}^{N}M_{il}V_{lj}-\sum_{q=1}^{N}V_{iq}M_{jq}+V_{ij}\\
&=\sum_{q=1}^{N}(\delta_{jq}-M_{jq})V_{iq}+\sum_{l=1}^{N}(\delta_{il}-M_{il})V_{lj}.
\end{aligned}$$

Finally, with the definition of $\mathbf{Q}(\tau)$ in Eq. (10) and Eqs. (5a), (5b) follows

$$\begin{aligned}
&m_{ij}^{(2)}(\mathbf{x},\tau)\\
&=\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{V})*m_{ij}^{(2)}(\mathbf{x},\tau)\\
&\quad+\big(Q_{ij}(\tau)+Q_{ji}(\tau)\big)\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{V})*m^{(0)}(\mathbf{x})\\
&\quad+\sum_{k=1}^{N}Q_{jk}\frac{\partial}{\partial x_k}\Big(\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{V})*m_i^{(1)}(\mathbf{x},\tau)\Big)\\
&\quad+\sum_{k=1}^{N}Q_{ik}\frac{\partial}{\partial x_k}\Big(\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{V})*m_j^{(1)}(\mathbf{x},\tau)\Big)\\
&\quad+\sum_{k=1}^{N}\sum_{m=1}^{N}Q_{jm}Q_{ik}\frac{\partial}{\partial x_k}\frac{\partial}{\partial x_m}\Big(\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{V})*m^{(0)}(\mathbf{x})\Big)\\
&=\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{V})*m_{ij}^{(2)}(\mathbf{x},\tau)+\big(Q_{ij}(\tau)+Q_{ji}(\tau)\big)m^{*(0)}(\mathbf{x})\\
&\quad+\sum_{k=1}^{N}Q_{jk}\frac{\partial}{\partial x_k}\Big(\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{V})*m_i^{(1)}(\mathbf{x},\tau)\Big)\\
&\quad+\sum_{k=1}^{N}Q_{ik}\frac{\partial}{\partial x_k}\Big(\mathcal{G}(\mathbf{x},\mathbf{0},\mathbf{V})*m_j^{(1)}(\mathbf{x},\tau)\Big)\\
&\quad+\sum_{k=1}^{N}\sum_{m=1}^{N}Q_{jm}Q_{ik}\frac{\partial}{\partial x_k}\frac{\partial}{\partial x_m}m^{*(0)}(\mathbf{x})\\
&=\rho_Y(\mathbf{x})*m_{ij}^{(2)}(\mathbf{x},\tau)+\big(Q_{ij}(\tau)+Q_{ji}(\tau)\big)m^{*(0)}(\mathbf{x})\\
&\quad+\sum_{k=1}^{N}Q_{jk}\frac{\partial}{\partial x_k}\left(m_i^{*(1)}(\mathbf{x},\tau)-\sum_{r=1}^{N}Q_{ir}\frac{\partial}{\partial x_r}m^{*(0)}(\mathbf{x})\right)\\
&\quad+\sum_{k=1}^{N}Q_{ik}\frac{\partial}{\partial x_k}\left(m_j^{*(1)}(\mathbf{x},\tau)-\sum_{s=1}^{N}Q_{js}\frac{\partial}{\partial x_s}m^{*(0)}(\mathbf{x})\right)\\
&\quad+\sum_{k=1}^{N}\sum_{m=1}^{N}Q_{jm}Q_{ik}\frac{\partial}{\partial x_k}\frac{\partial}{\partial x_m}m^{*(0)}(\mathbf{x})\\
&=\rho_Y(\mathbf{x})*m_{ij}^{(2)}(\mathbf{x},\tau)+\Bigg(Q_{ij}(\tau)+Q_{ji}(\tau)\\
&\quad-\sum_{k=1}^{N}\sum_{m=1}^{N}Q_{jm}Q_{ik}\frac{\partial}{\partial x_k}\frac{\partial}{\partial x_m}\Bigg)m^{*(0)}(\mathbf{x})\\
&\quad+\sum_{k=1}^{N}Q_{jk}\frac{\partial}{\partial x_k}m_i^{*(1)}(\mathbf{x},\tau)+\sum_{k=1}^{N}Q_{ik}\frac{\partial}{\partial x_k}m_j^{*(1)}(\mathbf{x},\tau).
\end{aligned} \tag{A.10}$$

Appendix B. Auxiliary integral in Eq. (13)

To prove the equality in Eq. (15) we start by stating that, through integration by parts over one single variable x_i:

$$\begin{aligned}
&\int_{x_i}\left(\frac{\partial}{\partial x_i}m^{*(0)}(\mathbf{x})\right)x_j dx_i\\
&=m^{*(0)}(\mathbf{x})x_j\Big|_{x_i=-\infty}^{x_i=\infty}-\int_{x_i}\left(\frac{\partial}{\partial x_i}x_j\right)m^{*(0)}(\mathbf{x})dx_i\\
&=m^{*(0)}(\mathbf{x})x_j\Big|_{x_i=-\infty}^{x_i=\infty}-\delta_{ij}\int_{x_i}m^{*(0)}(\mathbf{x})dx_i
\end{aligned} \tag{B.1}$$

and using the fact that $\lim_{x_i\to\infty}m^{*(0)}(\mathbf{x})x_j-\lim_{x_i\to-\infty}m^{*(0)}(\mathbf{x})x_j = 0\ \forall i,j$ as well as the relation $\int_{\mathbf{x}}m^{*(0)}(\mathbf{x})d\mathbf{x}=1$ yields Eq. (15)

$$\begin{aligned}
&\int_{\mathbf{x}}x_jH_i^{(1)}(\mathbf{A},\mathbf{B},m^{*(0)}(\mathbf{x}))d\mathbf{x}\\
&=\sum_{k=1}^{N}Q_{ik}(\tau)\int_{\mathbf{x}}x_j\frac{\partial}{\partial x_k}m^{*(0)}(\mathbf{x})d\mathbf{x}\\
&=\sum_{k=1}^{N}Q_{ik}(\tau)\int_{x_1}\cdots\int_{x_{k-1}}\int_{x_{k+1}}\cdots\int_{x_N}m^{*(0)}(\mathbf{x})x_j dx_1\ldots dx_{k-1}dx_{k+1}dx_N\\
&\quad-\delta_{kj}\int_{\mathbf{x}}m^{*(0)}(\mathbf{x})d\mathbf{x}\\
&=-\sum_{k=1}^{N}Q_{ik}(\tau)\delta_{kj}=-Q_{ij}(\tau).
\end{aligned} \tag{B.2}$$

Please cite this article in press as: T. Scholz et al., Parameter-free resolution of the superposition of stochastic signals, Phys. Lett. A (2016), http://dx.doi.org/10.1016/j.physleta.2016.09.057

Appendix C. Computation of parameter matrix B from covariance matrix V

The measurement noise parameter **B** can be obtained from Eq. (11) through integration by parts:

$$\begin{aligned}\mathbf{V} &= \int_0^\infty e^{-\mathbf{A}s}\mathbf{B}e^{-\mathbf{A}^T s}ds \\ &= -\left[e^{-\mathbf{A}s}\mathbf{B}e^{-\mathbf{A}^T s}(\mathbf{A}^T)^{-1}\right]_0^\infty - \int_0^\infty \mathbf{A}e^{-\mathbf{A}s}\mathbf{B}e^{-\mathbf{A}^T s}(\mathbf{A}^T)^{-1}ds \\ &= -\lim_{s\to\infty}\left[e^{-\mathbf{A}s}\mathbf{B}e^{-\mathbf{A}^T s}(\mathbf{A}^T)^{-1}\right] + \mathbf{B}(\mathbf{A}^T)^{-1} \\ &\quad - \left[\mathbf{A}\int_0^\infty e^{-\mathbf{A}s}\mathbf{B}e^{-\mathbf{A}^T s}ds(\mathbf{A}^T)^{-1}\right] \\ &= \mathbf{B}(\mathbf{A}^T)^{-1} - \mathbf{A}\mathbf{V}(\mathbf{A}^T)^{-1} \end{aligned} \tag{C.1}$$

References

[1] J. Albersmeyer, M. Diehl, The lifted newton method and its application in optimization, SIAM J. Optim. 20 (3) (2010) 1655–1684.
[2] J. Andersson, B. Houska, Towards a computer algebra system with automatic differentiation for use with object-oriented modelling languages, in: Object-Oriented Modeling Languages, 2010.
[3] F. Böttcher, J. Peinke, D. Kleinhans, R. Friedrich, P.G. Lind, M. Haase, Reconstruction of complex dynamical systems affected by strong measurement noise, Phys. Rev. Lett. 97 (Sep. 2006) 090603.
[4] R. Friedrich, J. Peinke, Description of a turbulent cascade by a Fokker–Planck equation, Phys. Rev. Lett. 78 (Feb. 1997) 863–866.
[5] R. Friedrich, J. Peinke, M. Sahimi, M.R.R. Tabar, Approaching complexity by stochastic methods: from biological systems to turbulence, Phys. Rep. 506 (5) (2011) 87–162.
[6] C. Gardiner, Stochastic Methods, Springer-Verlag, Berlin, Heidelberg, 2009.
[7] B. Lehle, Analysis of stochastic time series in the presence of strong measurement noise, Phys. Rev. E 83 (Feb. 2011) 021113.
[8] B. Lehle, Stochastic time series with strong, correlated measurement noise: Markov analysis in n dimensions, J. Stat. Phys. 152 (6) (2013) 1145–1169.
[9] P.G. Lind, M. Haase, F. Böttcher, J. Peinke, D. Kleinhans, R. Friedrich, Extracting strong measurement noise from stochastic time series: applications to empirical data, Phys. Rev. E 81 (Apr. 2010) 041125.
[10] P.G. Lind, A. Mora, J.A.C. Gallas, M. Haase, Reducing stochasticity in the north Atlantic oscillation index with coupled Langevin equations, Phys. Rev. E 72 (Nov. 2005) 056706.
[11] M. Luhur, J. Peinke, J. Schneemann, M. Wächter, Stochastic modeling of lift and drag dynamics under turbulent wind inflow conditions, Wind Energy 18 (2014) 317–337.
[12] A. Martynyuk, Stability by Liapunov's Matrix Function Method with Applications, CRC Press, 1998.
[13] P. Milan, M. Wächter, J. Peinke, Turbulent character of wind energy, Phys. Rev. Lett. 110 (Mar. 2013) 138701.
[14] B. Prusseit, K. Lehnertz, Stochastic qualifiers of epileptic brain dynamics, Phys. Rev. Lett. 98 (2007) 138103.
[15] F. Raischel, T. Scholz, V.V. Lopes, P.G. Lind, Uncovering wind turbine properties through two-dimensional stochastic modeling of wind dynamics, Phys. Rev. E 88 (Oct. 2013) 042146.
[16] C. Renner, J. Peinke, R. Friedrich, Evidence of Markov properties of high frequency exchange rate data, Physica A 298 (3–4) (2001) 499–520.
[17] C. Renner, J. Peinke, R. Friedrich, Experimental indications for Markov properties of small-scale turbulence, J. Fluid Mech. 433 (2001) 383–409.
[18] A. Wächter, L.T. Biegler, On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming, Math. Program. 106 (2006) 25–57.
[19] V. Zaburdaev, S. Uppaluri, T. Pfohl, M. Engstler, R. Friedrich, H. Stark, Langevin dynamics deciphers the motility pattern of swimming parasites, Phys. Rev. Lett. 106 (May 2011) 208103.
[20] D.V. Lukyanenko, Zhang Ye, A.G. Yagola, An optimal regularization method for convolution equations on the sourcewise represented set, J. Inverse Ill-Posed Probl. (2015).
[21] F. Zhang, The Schur Complement and Its Applications, Numerical Methods and Algorithms, vol. 4, Springer US, 2005.

Please cite this article in press as: T. Scholz et al., Parameter-free resolution of the superposition of stochastic signals, Phys. Lett. A (2016), http://dx.doi.org/10.1016/j.physleta.2016.09.057

Chapter 8

Principal wind turbines for a conditional portfolio approach to wind farms

One of the main motivations of applying stochastic methods to wind energy production data as shown in the previous chapters is its fluctuating nature. This intermittency raises several issues, for example considering bidding on the liberalized electricity market or reserve commitment.
This chapter explores, how data measured in a wind park can be used to quantify the risk associated to wind power production, which can then, in turn be used to devise bidding or reserve commitment strategies. The method introduced in this chapter proposes a conditional risk-return approach to incorporate knowledge of the wind field into the measure. The main challenge here is the volume of the data collected from all turbines in a wind park. For this reason, a principal component analysis is performed on wind speed and direction data and based on that, a wind park "state" is derived. The proposed measure is based on the risk and return conditioned to this state of the wind park [a].

[a] Equation (7) in this paper reads

$$S^{(q)}(t) = \frac{\sum_{i=1}^{q} \left| \lambda_i \sum_{j=1}^{N_W} \sum_{k=1}^{N_\tau} \omega_l V_j(t-\tau_k) \right|}{\sum_{i=1}^{q} |\lambda_i|},$$

where $l = (i-1)N_W + j$.

Principal wind turbines for a conditional portfolio approach to wind farms

2014 J. Phys.: Conf. Ser. 524 012183

(http://iopscience.iop.org/1742-6596/524/1/012183)

The Science of Making Torque from Wind 2014 (TORQUE 2014) IOP Publishing
Journal of Physics: Conference Series **524** (2014) 012183 doi:10.1088/1742-6596/524/1/012183

Principal wind turbines for a conditional portfolio approach to wind farms

Vitor V. Lopes[1,2], **Teresa Scholz**[3,4], **Frank Raischel**[5], **Pedro G. Lind**[6]

[1]DEIO-CIO, Science faculty, University of Lisbon, Portugal
[2]Universidad de las Fuerzas Armadas-ESPE, Latacunga, Ecuador
[3]Center for Theoretical and Computational Physics, University of Lisbon, Portugal
[4]Energy Analysis and Networks Unit, National Laboratory of Energy and Geology, Lisbon, Portugal
[5]Center for Geophysics, IDL, University of Lisbon, Portugal
[6]ForWind-Center for Wind Energy Research, Institute of Physics, Carl-von-Ossietzky University of Oldenburg, Oldenburg, Germany

E-mail: teresa.scholz@lneg.pt

Abstract. We introduce a measure for estimating the best risk-return relation of power production in wind farms within a given time-lag, conditioned to the velocity field. The velocity field is represented by a scalar that weighs the influence of the velocity at each wind turbine at present and previous time-steps for the present "state" of the wind field. The scalar measure introduced is a linear combination of the few turbines, that most influence the overall power production. This quantity is then used as the condition for computing a conditional expected return and corresponding risk associated to the future total power output.

1. Introduction

Wind energy is becoming a top contributor to the renewable energy mix due to rather high capacities and generation costs that are becoming competitive with conventional energy sources[1]. However, wind energy systems suffer from a major drawback, the fluctuating nature of their source[2], which imposes a challenge to the wind power producer when it comes to trading on the liberalized electricity markets. For that, participants must bid in advance and the uncertainty of wind power production can lead to differences between the committed and actually produced energy [3]. This imbalance may result in the payment of penalties which decreases the revenue [4]. Therefore, to achieve maximum profit, it is necessary to develop optimal offering strategies. A review on the methods employed for deriving bidding strategies is given by[5].

The uncertainty associated to wind energy also affects the wind park and power grid operators, especially concerning reserve committment. An appropriate reserve committment needs to be a trade-off between economic (additional capacity costs) and reliability issues (risk of loss of load) [6]. The operators therefore need to develop a risk managing scheme, taking into account the intermittency of the energy source [7]. A review on the economic dispatch and risk management considering wind power in the power market is given by [8].

In this study we propose a measure to asses the return of a wind farm in terms of risk that can be used to develop bidding strategies for market participation. Our measure is based on the mean-variance portfolio (MVP) or risk-return approach first introduced in the context of portfolio selection by Markowitz[9]. Markowitz defines the return on a portfolio as a weighted sum of random variables where the investor can choose the weights. For investment he proposes to assess the ratio between the expected

The Science of Making Torque from Wind 2014 (TORQUE 2014) IOP Publishing
Journal of Physics: Conference Series **524** (2014) 012183 doi:10.1088/1742-6596/524/1/012183

return and the associated risk, which he defines as the standard deviation of return. An investment decision then is a trade-off between risk and expected return, since "the portfolio with maximum expected return is not necessarily the one with minimum variance"[9]. However, for a given amount of risk, MVP allows to maximize the expected return for a given level of risk or equivalently to minimize the risk for a given level of expected return.

In the context of wind energy this translates into optimizing the trade-off between maximizing wind power output and minimizing its variability. MVP has been employed in the framework of wind energy by Roques et al[10] to define optimal cross-countries wind power portfolios. Using historical wind production data from five European countries the authors attempt two case studies. First, they optimize the wind power output and, second, maximize the wind power contribution to system reliability. The risk-return approach has also been employed by Kitzing[11] with the purpose to assess the risk implications of two support instruments, feed-in tariffs and feed-in premiums.

This paper introduces a new approach of risk-return evaluation by including a dependence on the state of the full wind farm. More precisely, our risk-return approach is *conditioned* to the wind field at each time-step. It is known that such conditional stochastic approaches allow to tackle the non-stationary character of wind[2, 12].

The paper is organized as follows. Section 2 describes the used datasets. In Sec. 3 the employed methodology is introduced, first, the conditional risk-return quotient based on this scalar is defined and then a scalar observable is derived for quantifying the velocity field taken at specific time instants. In Sec. 4 we present and explain the results obtained for a wind farm in Portugal and Sec. 5 concludes the paper.

2. Data: The wind field and total power production

The data analyzed in this manuscript were obtained from a wind park with 80 turbines located in a mountainous region of Portugal. The data comprise three years of measurements with a ten-minute sampling time (1.7×10^{-3} Hz). They consist of two sets, the wind velocity field observed at each wind energy converter (WEC) in the farm and the farm's total power production. It should be noted that the wind speed is measured by a nacelle anemometer on each WEC and therefore has some uncertainty associated to it. An equivalent wind speed which is affecting the rotor disk could be derived (see e.g. Ref.[13]) and used for further analysis. Since previous research successfully used nacelle anemometer measurements for a similar purpose, namely for reproducing the power output generated by single turbines and wind farms[14], we also use such nacelle measurements. The velocity field is represented as a matrix $\mathbf{V} \in \mathbb{C}^{N_W \times N_T}$, where each entry $V_n(t) = V_n^{(x)}(t) + iV_n^{(y)}(t)$ corresponds to the velocity vector with x- and y-components $V_n^{(x)}(t)$ and $V_n^{(y)}(t)$ at the WEC labelled as $n = 1, \ldots, N_W$ and time $t = 1, \ldots, N_T$:

$$\mathbf{V} = \begin{bmatrix} V_1(1) & V_1(2) & V_1(3) & \ldots & V_1(N_T) \\ V_2(1) & V_2(2) & V_2(3) & \ldots & V_2(N_T) \\ V_3(1) & V_3(2) & V_3(3) & \ldots & V_3(N_T) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ V_{N_W}(1) & V_{N_W}(2) & V_{N_W}(3) & \ldots & V_{N_W}(N_T) \end{bmatrix}, \tag{1}$$

where $V_i(t)$ is the velocity at turbine i at time-step t. Vertical velocities are always neglected.

Our aim is to estimate the expected power production $P(t+\tau)$ at a time-lag τ after the present time t, using the velocity field observations at time t as input. The shape of the wind power curve suggests that fluctuations in wind speed above the rated wind speed should not result in large wind power fluctuations for a single turbine. However, we find significant power fluctiations in this regime and therefore do not neglect those measurements. The methodology is described in the next section.

The Science of Making Torque from Wind 2014 (TORQUE 2014) IOP Publishing
Journal of Physics: Conference Series **524** (2014) 012183 doi:10.1088/1742-6596/524/1/012183

3. Methodology and data analysis

This section describes the methodology used for the data analysis. A new conditional risk-return approach to power production in a wind farm is introduced. It is stressed that any risk measure should take into account the present "state" of the wind farm. The straightforward way to define this state is through the full set of wind velocities at a particular time. However, as we stress below, the full wind field would imply a large amount of data to determine the state of the system. Therefore, instead of the full velocity field, we define a scalar quantity that is also capable to characterize the state of a spatially extended system such as a wind farm.

The usual way to define the percentage power return $r(t)$ is[11]

$$r(t) = \frac{P(t+\Delta t) - P(t)}{P(t)}, \tag{2}$$

where Δt is a fixed time-lag. Having the time series of the returns we can then define the expected power return by

$$\hat{r} = \int_{-\infty}^{\infty} r\rho(r(t))dr, \tag{3}$$

and the associated risk is given by the variance of the expected returns, namely

$$\Delta r = \int_{-\infty}^{\infty} (r-\hat{r})^2\rho(r(t))dr, \tag{4}$$

where $\rho(r(t))$ is the probability for having a return $r(t)$ at time t.

In both Eqs. (3) and (4) the return r considers the total power output in the wind farm. Alternative choices are possible, e.g. to account only for the power of a subset of all WECs, the ones that represent the most the power output in the wind farm. An improvement to the power sum of the total or a subset of WECs would be a weighted sum, but in all cases it would yield an expected value and a variance (risk) independent of the present wind velocity field.

We propose instead to consider a proper quantity, "quantifying" the wind velocity field, and use it as condition for computing the expected return and corresponding variance. This state of one wind farm could be defined as the set of values of the wind velocities at instant t. However, since wind farms typically contain $N_W \sim 100$ wind turbines, each one with $N_s \sim 50$ admissible wind speed states after proper binning[15, 16], the phase space for the wind farm would comprehend approximately 50^{100} possible states.

To overcome this shortcoming, we choose a weighted sum based on the principal component analysis (PCA) of the wind speed at all the turbines and several time-lags. This weighted sum is truncated at a given order q, and we symbolize it henceforth as $S^{(q)}(t)$ (defined in Eq. (7)), which we introduce as follows.

The PCA is performed through the eigenvalue decomposition of a matrix constructed from the covariance matrix of wind velocities. Given a set of time-lags $\mathcal{T} = \{\tau_1, ..., \tau_{N_\tau}\}$, for each pair of turbines i and j, we compute the covariance associated to their wind velocities referenced to given time-lags $\tau_k, \tau_l \in \mathcal{T}$ by

$$\begin{aligned}
\mathbf{M}_{(i\tau_k)(j\tau_l)} &= \{C_{ij(\tau_l-\tau_k)}\}_{i,j=1,...,N_W} \\
&= \{\langle (V_i(t+\tau_k) - \langle V_i(t+\tau_k)\rangle)(V_j(t+\tau_l) - \langle V_j(t+\tau_l)\rangle)^*\rangle\}_{i,j=1,...,N_W} \\
&= \begin{bmatrix} M_{(1\tau_k)(1\tau_l)} & M_{(1\tau_k)(2\tau_l)} & \cdots & M_{(1\tau_k)(N_W\tau_l)} \\ M_{(2\tau_k)(1\tau_l)} & M_{(2\tau_k)(2\tau_l)} & \cdots & M_{(2\tau_k)(N_W\tau_l)} \\ \vdots & \vdots & \ddots & \vdots \\ M_{(N_W\tau_k)(1\tau_l)} & M_{(N_W\tau_k)(2\tau_l)} & \cdots & M_{(N_W\tau_k)(N_W\tau_l)} \end{bmatrix},
\end{aligned} \tag{5}$$

The Science of Making Torque from Wind 2014 (TORQUE 2014) IOP Publishing
Journal of Physics: Conference Series **524** (2014) 012183 doi:10.1088/1742-6596/524/1/012183

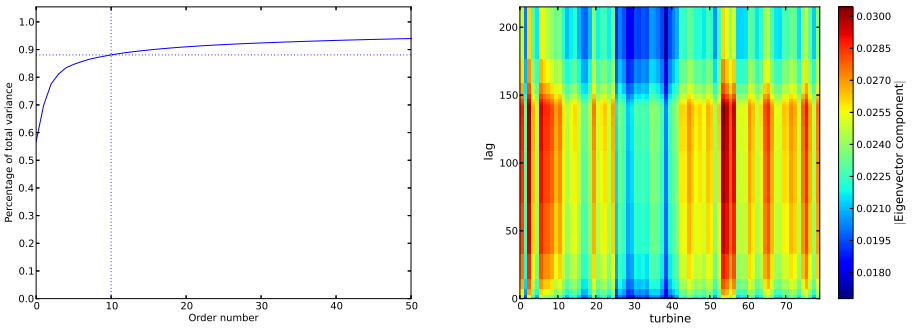


Figure 1: Percentage of total variance expressed by the first 50 eigenvectors (left) and the absolute value of the components of the eigenvector corresponding to the first eigenvalue (right) of the master matrix M.

where $i, j = 1, \ldots, N_W$, $l, k = 1, \ldots, N_\tau$ and $\langle\rangle$ denotes the average over time t and * denotes the complex conjugate.

Taking $\tau_\ell, \tau_k = 1, \ldots, N_\tau$ the covariances can be embedded in a master matrix defined as

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{(i0)(j0)} & \mathbf{M}_{(i0)(j1)} & \cdots & \mathbf{M}_{(i0)(jN_\tau)} \\ \mathbf{M}_{(i1)(j0)} & \mathbf{M}_{(i1)(j1)} & \cdots & \mathbf{M}_{(i1)(jN_\tau)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{M}_{(iN_\tau)(j0)} & \mathbf{M}_{(iN_\tau)(j1)} & \cdots & \mathbf{M}_{(iN_\tau)(jN_\tau)} \end{bmatrix} \tag{6}$$

which is a symmetric matrix of dimension $N_W N_\tau \times N_W N_\tau$.

The master-matrix has two important properties. The first one is the symmetry $M_{(i\tau_k)(j\tau_l)} = M_{(j\tau_l)(i\tau_k)}$. Notice that $\langle V_i(t+\tau_k)\rangle = \frac{1}{N_T-\tau_k}\sum_{t=\tau_k+1}^{N_T} V_i(t)$. Therefore, for sufficiently large number of measures, $N \gg \tau_i$, $\forall i$, we take $\langle V_i(t+\tau_k)\rangle = \langle V_i(t)\rangle$, $\forall i$. Thus, $\mathbf{M}_{(i\tau_k)(j\tau_l)} = \mathbf{M}_{(i0)(j(\tau_l-\tau_k))} = \mathbf{M}_{(j0)(i(\tau_k-\tau_l))} = \mathbf{M}^T_{(i0)(j(\tau_k-\tau_l))} = \mathbf{M}^T_{(i\tau_l)(j\tau_k)}$. The master-matrix $\mathbf{M}$ is therefore symmetric. Notice however that it can have complex eigenvalues and eigenvectors since $V_i(t+\tau_k)$ is complex $\forall i, k$. The second property is that, assuming approximation $\langle V_i(t+\tau_k)\rangle = \langle V_i(t)\rangle$, the master-matrix is Toeplitz by blocks.

The diagonalization of this covariance matrix allows to compute principal directions which form the set of components uncorrelated with each other. These directions in phase space are defined by linear combinations of the wind speed at each single turbine and different time-lags. The principal directions corresponding to the largest eigenvalues are the ones along which the system fluctuates strongly. Therefore, selecting from these linear combinations the few ones corresponding to the largest eigenvalues gives a subset of the most "influent" turbines for the global wind velocity state of the wind farm.

Figure 1 on the left shows the percentage of the total variance that is expressed by the first m (largest) eigenvalues of the master matrix. It can be seen that the ten largest eigenvalues already account for almost 90% of the total variance in the data (check dotted lines in Fig. 1). These facts will be important below to interpret our results.

Further the principal direction corresponding to the largest eigenvalue alone comprehends already almost 50% of signal's variance. Its components are shown on the right of Fig. 1. It can be concluded that the contribution of the turbines is not equal throughout the park, for instance turbines 54 to 57 have a high influence, turbines 25 to 42 a comparatively low one. Considering the contributions of the time-lags, i.e. the influence of information from the past, Fig. 1 shows that the contribution starts out small for zero

The Science of Making Torque from Wind 2014 (TORQUE 2014) IOP Publishing
Journal of Physics: Conference Series **524** (2014) 012183 doi:10.1088/1742-6596/524/1/012183

time lag and quickly reaches a plateau. It then decays rapidly for lags larger than one day. This pattern is especially prominent for turbines with lower contributions.

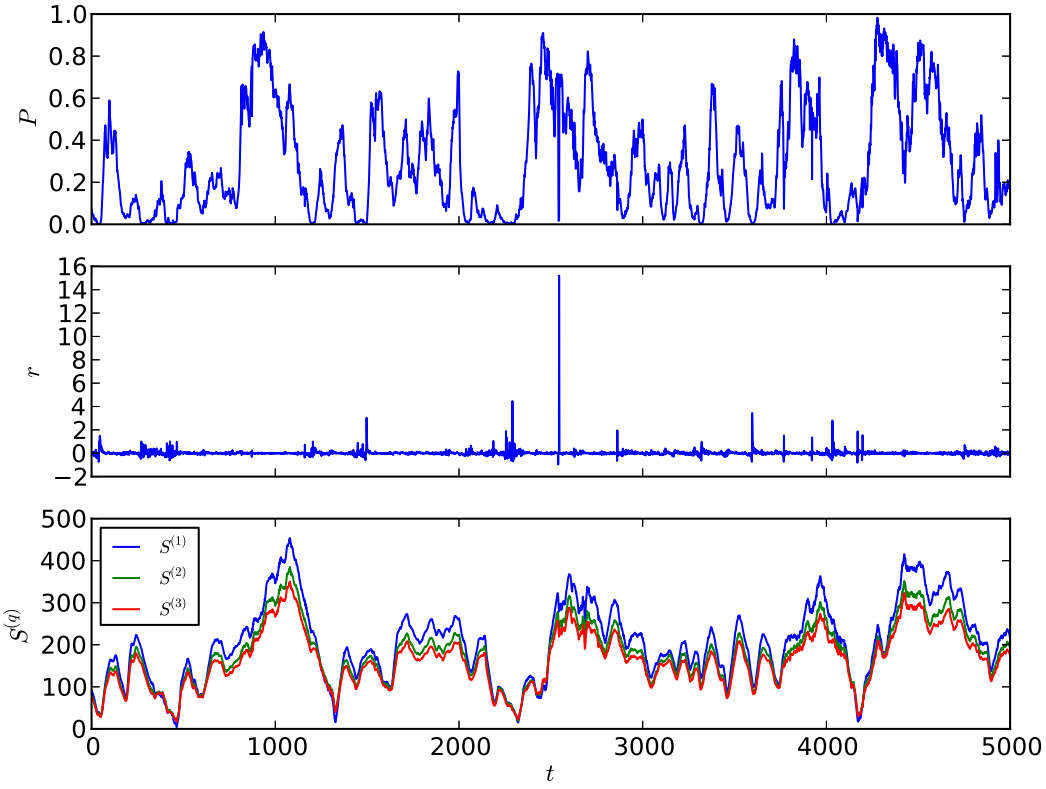


Figure 2: Subset of the time series of the wind park's total power output (top), return (middle) and wind park state $S^{(q)}$ for $q = 1, 2, 3$ (bottom).

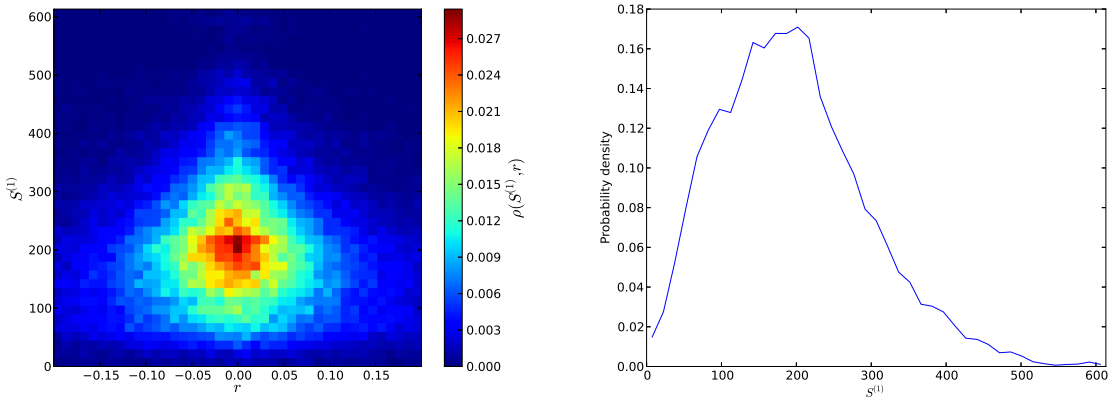


Figure 3: Joint probability density function of the wind park state $S^{(1)}$ and the power return r (left) and probability density function of the wind park state $S^{(1)}$ (right).

The Science of Making Torque from Wind 2014 (TORQUE 2014) IOP Publishing
Journal of Physics: Conference Series **524** (2014) 012183 doi:10.1088/1742-6596/524/1/012183

Next, we use the insight provided by a set of eigenvectors ordered according to the magnitude of their corresponding eigenvalue to define a scalar quantity for characterizing the state of the wind farm. Namely, we consider the first q eigenvalues $\lambda_1, \ldots, \lambda_q$ of $\mathbf{M}$ with $q = 1, \ldots, N_W N_\tau$, together with their corresponding eigenvectors and define the scalar quantity

$$S^{(q)}(t) = \frac{\sum_{i=1}^{q} \left| \lambda_i \sum_{j=1}^{N_W} \sum_{k=1}^{N_\tau} \omega_{ij} V_j(t - \tau_k) \right|}{\sum_{i=1}^{q} |\lambda_i|}, \tag{7}$$

where ω_{ij} describes the j-th component of the eigenvector to the i-th eigenvalue λ_i and τ_k is given in units of the time increment between successive measurements of the wind speed.

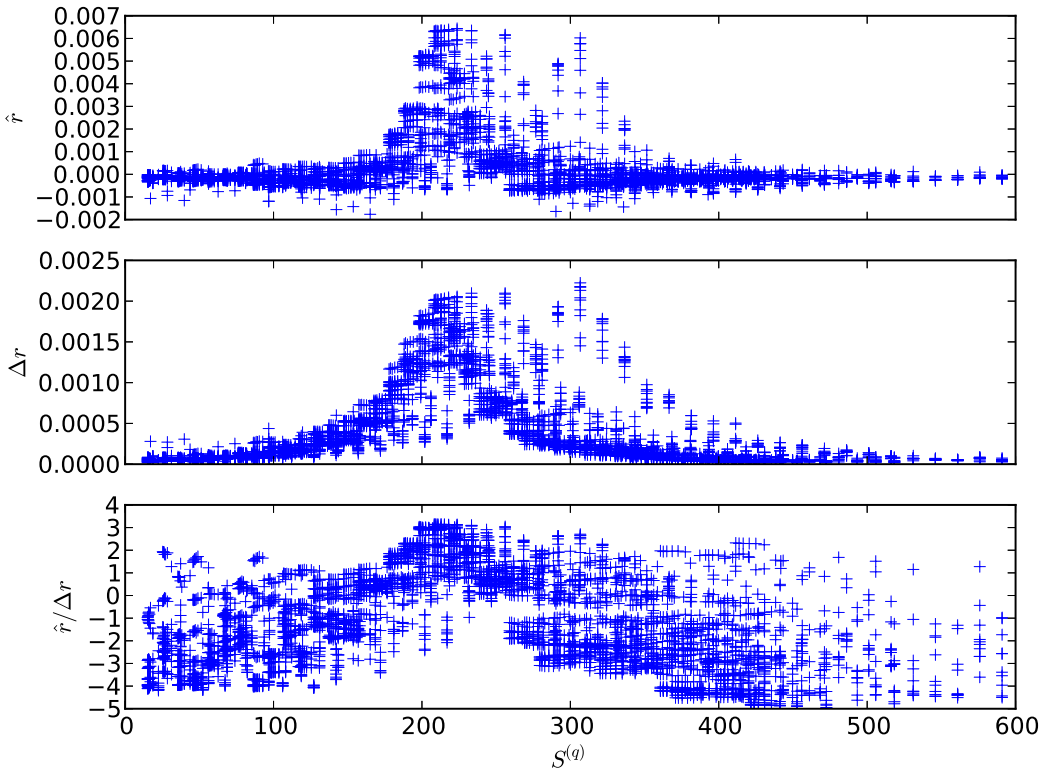


Figure 4: Conditional returns $\hat{r}(q, \tau, S^{(q)}(t))$ (top), risks $\Delta r(q, \tau, S^{(q)}(t))$ (middle) and risk-return quotient $\hat{r}(q, \tau, S^{(q)}(t))/\Delta r(q, \tau, S^{(q)}(t))$ (bottom) as a function of the state $S^{(q)}$ for several values of $q = 1, \ldots, 10$ and $\tau = 1, \ldots, 10$.

We argue here that $S^{(q)}(t)$ quantifies the wind velocity state of the wind farm at time t. Since the eigenvalues are ordered the first sum in Eq. (7) comprehends the q largest eigenvalues. The other two sums are in the set of WECs and in (previous) times. Thus, for each choice of WEC j and time-delay τ_k the quantity S uses a weight given by the j-th component of the i-th most influent eigenvector. In other words, by weighting a sum of velocity measurements across the wind farm and, simultaneously, at different time-steps, the quantity $S^{(q)}$ incorporates both spatial and temporal information of the wind speed observed at the wind farm. We call $S^{(q)}$ the state of the wind farm.

Figure 2 (bottom) shows the evolution of $S^{(q)}$ for $q = 1, 2$ and 3 together with the time series of the wind park's total power output (top) and the corresponding returns given by Eq. (2) (middle). One can

The Science of Making Torque from Wind 2014 (TORQUE 2014) IOP Publishing
Journal of Physics: Conference Series **524** (2014) 012183 doi:10.1088/1742-6596/524/1/012183

observe that adding information from the third largest eigenvalue and its corresponding eigenvector leads to very small changes in $S^{(q)}$, especially when $S^{(q)}$ is small. Similarly to what was already concluded when addressing Fig. 1, such similar evolutions indicate that the lowest q values already capture most of the information in the velocity field, at least relatively to its energy content.

Using $S^{(q)}$, we introduce a *conditional* risk-return approach, where expected return and risk are given respectively by

$$\hat{r}(q,\tau,S^{(q)}(t)) = \int_{-\infty}^{\infty} r\rho(r(t+\tau)|S^{(q)}(t))dr \tag{8}$$

and

$$\Delta r(q,\tau,S^{(q)}(t)) = \int_{-\infty}^{\infty} (r-\hat{r})^2\rho(r(t+\tau)|S^{(q)}(t))dr. \tag{9}$$

The joint probability density function $\rho(r(t),S^{(1)})$ for both the state $S^{(1)}(t)$ and the total power output return $r(t)$ is shown in Fig. 3 (left) together with the marginal probability density function for $S^{(1)}$ (right). The joint distribution is approximately symmetric around $r=0$ and its variance decreases with the value of S, i.e. the expected variability of the power decreases with the higher intensity of the wind field.

Both density functions in Fig. 3 are needed for deriving the conditional probability $\rho(r(t+\tau)|S^{(q)}(t)) = \rho(r(t+\tau),S^{(q)}(t))/\rho(S^{(q)}(t))$ needed for the computation of the conditional risk-return measure in Eqs. (8) and (9).

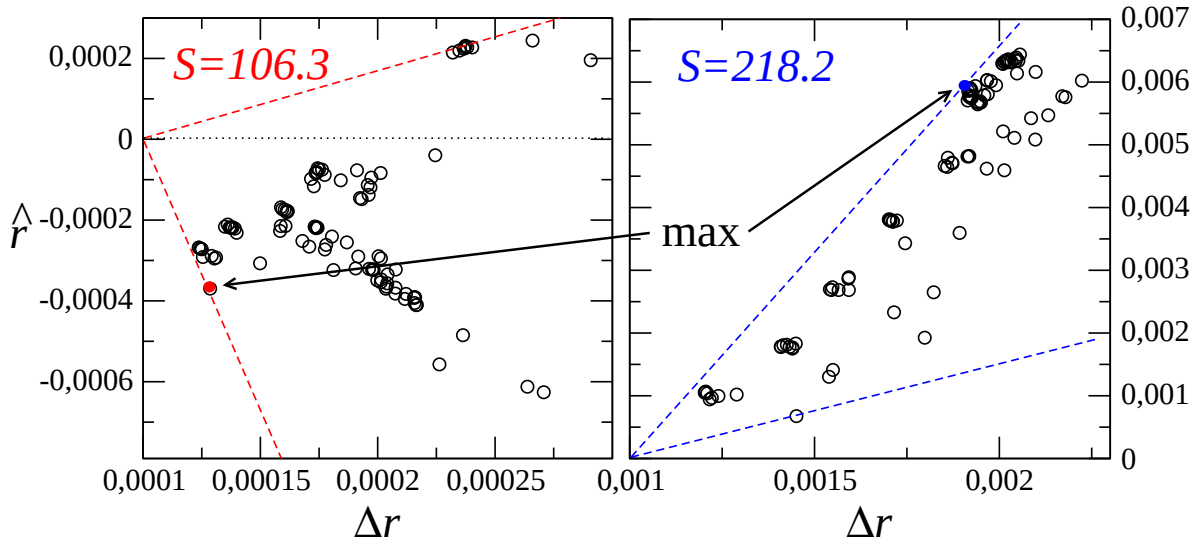


Figure 5: Risk-return diagrams for two different values of $S^{(q)}(t)$. Each circle corresponds to a pair (q, τ). For $S^{(q)}(t)=106.3$ the maximum of the risk-return quotient is obtained for $q_{max}=1$ and $\tau_{max}=1$ (red bullet), for $S^{(q)}=218.2$ however, one obtains the values $q_{max}=10$ and $\tau_{max}=4$ (blue bullet).

4. Results

Using Eqs. (8) and (9), we can now compute the conditional returns $\hat{r}(q,\tau,S^{(q)}(t))$, risks $\Delta r(q,\tau,S^{(q)}(t))$ and risk-return ratios $\hat{r}/\Delta r$ as a function of the state $S^q(t)$, i.e. fixing the values of q and τ. Results are shown in Fig. 4, where q and r range between 1 and 10. Important to notice here is that the conditional risk-return ratio varies a lot with q and τ within a wind farm state $S^{(q)}(t)$. Therefore, a possible optimal bidding strategy would be to take into account which values of q and τ yield the maximum ratio.

The Science of Making Torque from Wind 2014 (TORQUE 2014) IOP Publishing
Journal of Physics: Conference Series **524** (2014) 012183 doi:10.1088/1742-6596/524/1/012183

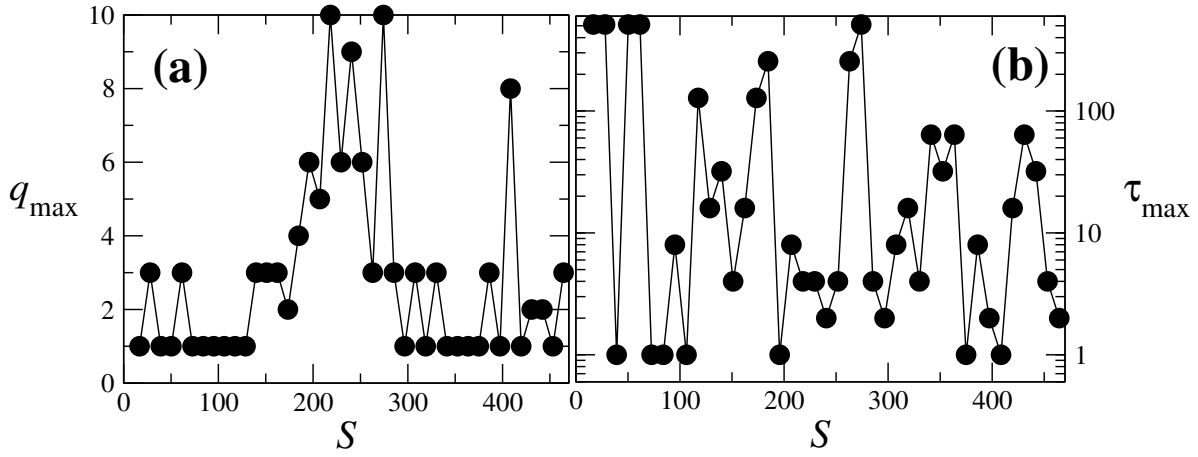


Figure 6: Values that should be selected for q_{max} and τ_{max} for each state of the wind farm.

This sensitivity of the optimal risk-return ration to the state of the wind farm is exemplified in detail in Fig. 5 for two particular values of $S^{(q)}(t)$: for $S^{(q)} = 106.3$ the maximum of the risk-return quotient is obtained for $q_{max} = 1$ and $\tau_{max} = 1$ (red bullet), for $S^{(q)} = 218.2$ however, one obtains the values $q_{max} = 10$ and $\tau_{max} = 4$ (blue bullet).

Figure 6 shows the time series of optimal values, q_{max} and τ_{max} for the full range of possible states observed in the wind farm analyzed in this study.

The values of q_{max} have values ranging from one up to ten. The lowest values - typically $q_{max} = 1, 2$ or 3 - are attained for very weak winds (small values of S, left side of the spectrum) or for wind gusts (large values of S, right side of the spectrum). This can be easily explained: the two extreme situations are the ones for which the wind turbines are more synchronized and therefore most of the variance is already included in very first eigen-modes, corresponding to the lowest values of q_{max}. In the middle range of the wind speed (and also of the state S) the wind farm is far more heterogeneous and therefore higher order eigen-modes are necessary to explain the variability of the power output, yielding larger values for q_{max}.

As for τ_{max} the variability is much stronger, ranging from one up to several hundreds. The values chosen for τ_{max} were the ones that numerically maximize the absolute value of the risk-return ratio. Whether for other values of τ the risk-return ratio for a given state S is *approximately* the same needs to be still investigated.

5. Conclusions and discussion

In this paper we extend the standard portfolio analysis introducing a maximum risk-return ratio conditioned to the present state of the system and apply it to one wind farm, by taking the observed wind velocity field into account. For each defined wind farm state, the corresponding optimal risk-return ratio yields a particular time-lag (τ) which gives the best time-horizon to make power output forecasts.

Our results provide evidence that it is sensible to select the prediction horizon depending on the present state of the wind speed field instead of fixing it, as it is standardly done. Since, as we know, fixing the time-lag for the forecast, independently of the present state of the wind field, can lead to the underestimation of the risk levels for a given expected return or to overestimation of the return level for a given expected risk.

We conjecture that this dependence on the state of the system happens due to the non-stationary and intermittent character of the wind velocity field[2, 12], not only at one WEC but at the level of the entire wind park. Therefore, such a procedure can be helpful in other situations dealing with non-stationary and intermittent sets of data, such as the ones commonly observed in brain research and finance.

The Science of Making Torque from Wind 2014 (TORQUE 2014) IOP Publishing
Journal of Physics: Conference Series **524** (2014) 012183 doi:10.1088/1742-6596/524/1/012183

Three important remarks raise from the conclusions of our results. First, the conditional portfolio approach can be taken to derive other more sophisticated risk measurements, such as the value-at-risk for the total power output[17].

Second, the time evolution of the state of the wind farm as defined above, while properly reflecting the physical situation of interest here, could be studied in deeper detail elsewhere, particularly in what concerns its (non)-stationary character and possible intermittency of its increments in different time-scales. Depending on the outcome a Langevin approach already successfully applied to power output and other WEC properties[2, 18], may provide further insight of the evolution of wind farms.

Third, as explained above, the wind speed measurments from the nacelle anemometer have a large uncertainty which may be causing the variations of the risk-return estimates in Fig. 6. Improvements in our analysis could be achieved by substituting the nacelle measurements by effective velocity values, which have less uncertainty. These and other issues will be addressed elsewhere.

Acknowledgments

The authors thank Philip Rin, João P. da Cruz for useful discussions and GENERG, SA. for providing the original data. The authors thank Fundação para a Ciência e a Tecnologia for financial support under PEst-OE/FIS/UI0618/2011, PEst-OE/MAT/UI0152/2011, FCOMP-01-0124-FEDER-016080, SFRH/BPD/65427/2009 (FR) and SFRH/BD/86934/2012 (TS). PGL thanks the German Environment Ministry and VVL the Prometeo Project of SENESCYT (Ecuador) for financial support (0325577B).

References

[1] J. Wen, Y. Zheng and F. Donghan, Renewable and Sustainable Energy Reviews **13**(9), 2485-2494, 2009.
[2] P. Milan, M. Wächter and J. Peinke, Phys. Rev. Lett. **110**, 138701, 2013.
[3] M. Hosseini-Firouz, International Journal of Electrical Power and Energy Systems **49**(0), 359-368, 2013.
[4] P.Pinson, C.Chevallier and G.N.Kariniotakis, Trading Wind Generation From Short-Term Probabilistic Forecasts of Wind Power, IEEE Transactions on Power Systems **22**(3), 1148-1156, 2007.
[5] G. Li, J. Shi and X. Qu, Energy **36**(8), 4686-4700, 2011.
[6] E.M. Gouveia and M.A. Matos, Electric Power Systems Research **79**(5), 734-739, 2009.
[7] H.M. Ghadikolaei, A. Ahmadi, J. Aghaei and M. Najafi, Renewable and Sustainable Energy Reviews **16**(7), 4734-4743, 2012.
[8] R. Boqiang and J. Chuanwen, Renewable and Sustainable Energy Reviews **13**(8), 2169-2174, 2009.
[9] H. Markovitz, The Journal of Finance **7**(1), 77-91, 1952.
[10] F. Roques, C. Hiroux and M. Saguan, Energy Policy **38**(7), 3245-3256, 2010.
[11] L. Kitzing, Energy **64**(0), 495-505, 2014.
[12] P.G. Lind, M. Wächter and J. Peinke, "Reconstructing the intermittent dynamics of the torque in wind turbines", submitted, 2014.
[13] E. Simley, L.Y. Pao, Design and evaluation of a wind speed estimator for hub-height and shear components, Proceedings of 9 th PhD Seminar on Wind Energy in Europe, Uppsala University Campus Gotland Sweden, September, 2013.
[14] P. Milan, M. Wächter and J. Peinke, Stochastic Modeling of Wind Power Production, Proceedings of EWEA, 2011.
[15] V.V. Lopes, T. Scholz, A. Estanqueiro, and A.Q. Novais, "On the use of Markov chain models for the analysis of wind power time series", Environment and Electrical Engineering (EEEIC), 2012 11th International Conference on, 770-775, 2012.
[16] T. Scholz, V.V. Lopes, A. Estanqueiro, Energy **67**(0), 557-568, 2014.
[17] J. Longerstaey and P. Zangari, *RiskMetrics Technical Document*, Morgan Guaranty Trust Company of New York, 4th Ed., 1996.
[18] F.Raischel, T.Scholz, V.V.Lopes and P.G.Lind, Phys. Rev. E **88**, 042146, 2013.

Chapter 9

Conclusions

The main topic of this thesis was the development of stochastic methods for the analysis and simulation of wind energy data. For this, the evolution of wind speed, power and direction was assumed to follow a stochastic process and modeled with nonlinear Langevin equations, extracted directly from the data.

This approach allowed to describe the dynamic behavior of the complex wind energy systems with only a few coefficients, namely the drift and the diffusion coefficients, representing the deterministic and stochastic behavior of the process, respectively. The analyzed data was synthesized from a Markov chain model, which incorporates three variables: wind power, speed and direction. The original data was measured over a period of three years at a wind turbine in Portugal. The joint discretization yielded a model that captures the interdependency of the three variables. To use the original data to full capacity, not only one- but also two- and three-step transitions were incorporated in the model. The presented model is capable of capturing the statistical properties of the original time-series, namely the long-term distributions or persistence.

The synthesized wind data has a low sampling rate of 10-minute averages, usual in wind farms and turbine monitoring. A disputed topic in the context of the Langevin analysis of measured data is the computation of the τ-limit in the estimation process of the Kramers-Moyal coefficients, especially when applying it to time-series with a low sampling rate. In this thesis, it could be shown, that despite the original data consisting of 10-minute averages, the application of the Langevin method gave rise to sensitive results. In particular, the analysis of the drift field allowed to detect the working regimes of the wind turbine as well as its characteristic rated wind speed and power. Moreover, a method to derive the turbine specific power curve based on both the drift and the diffusion coefficient was proposed.

In many data sets collected from wind energy conversion systems, daily patterns could be detected, reflecting the non stationary character of wind data. Common Markov models neglect these diurnal patterns, even though they can be of interest, especially regarding scheduling and reserve commitment. In this thesis, the introduced models were extended to capture the time-dependent behavior of the underlying stochastic process. To estimate the transition probabilities, a time-dependent Markov process model based on Bernstein polynomials was used. Special attention was payed to the convex formulation of the optimization problem to ensure a unique optimum. By comparison with empirical data, it was shown, that the model can represent the long-term statistics as well as the daily patterns of the original data. Based on the transition matrices obtained from this model, drift and diffusion coefficients were computed, which allowed to monitor the change in the dynamic behavior of the underlying stochastic process during the day.

Together with the already existent literature, all these contributions show that a stochastic description of wind energy systems is possible and provides valuable insights into the dynamics of the wind energy conversion systems. However, a critical practical issue when applying the mentioned methods is, that often the real world data-series are spoiled with measurement noise and as a consequence the estimated drift and diffusion coefficients are distorted. In this scope, a

parameter-free methodology to disentangle the measurement noise from the underlying stochastic process was introduced. As a result, the unspoiled drift and diffusion coefficients were obtained together with the measurement noise parameters, which is demonstrated on synthetic data. Detached from the specific problem of measurement noise, the method in general serves to separate two sets of independent stochastic processes, given that one can be modeled as a Langevin- and the other as an Ornstein-Uhlenbeck process.

The prominent stochasticity of wind energy production analyzed in this thesis poses a challenge to the wind power producers concerning reserve commitment and also bidding on the liberalized energy markets. In this framework a new tool for developing bidding strategies was suggested, namely a risk-return measure that is conditioned to a wind park "state". It is derived from a principal component analysis of the velocity field of the whole park with different time-lags.

The findings presented in this thesis give rise to several open questions.

In the analysis of the wind turbine dynamics in the stationary framework, different working regimes of the wind turbine and the turbines rated wind speed could be discovered in the drift field projected into the speed-power plane. This investigation could be extended to the cyclic model including daily production patterns to see, if the found fixed point areas exist at every time of the day and how the underlying dynamics change. It would also be interesting to see, how the dynamic behavior changes when taking the wind speed and power into account, i.e. investigating the change of the drift and diffusion coefficients in different points of the speed-power plane.

The time-dependent Markov model introduced in this thesis can be used to model any cyclic Markov process. It would be interesting to apply it to longer time-series to see, if it is possible to also capture seasonal patterns in wind energy production.

In general, the method proposed to disentangle measurement noise from the underlying stochastic process could be used in turbulence experiments in a wind tunnel as a data preprocessing tool. A concrete application would be the analysis of lift and drag coefficients that have been measured on a blade in the wind tunnel and that are known to be superposed by a signal stemming from unsteady aerodynamic effects [Luhur et al., 2014].

Another interesting extension of the work presented in this thesis would be to compare predicted power levels and their associated risk as given by the proposed risk-return measure with historical data. Moreover, it would be interesting to investigate which wind park states can be associated with a low or high risk-return quotient and compare the results from different sites.

The intermittency of the wind not only poses challenges for bidding on the liberalized energy market, but also to grid operators, who have to ensure a stable operating regime. It would be interesting to see if the tools developed in this thesis can be used to analyze the role of wind dynamics in the context of power grid stability and optimization Menck et al. [2014].

Bibliography

Anderson, T. W. and Goodman, L. A. (1957). Statistical inference about markov chains. *The Annals of Mathematical Statistics*, 28(1):89–110.

Arnesano, M., Carlucci, A., and Laforgia, D. (2012). Extension of portfolio theory application to energy planning problem – the italian case. *Energy*, 39(1):112 – 124.

Awerbuch, S. (2006). Portfolio-based electricity generation planning: Policy implications for renewables and energy security. *Mitigation and Adaptation Strategies for Global Change*, 11(3):693–710.

Awerbuch, S., Berger, M., Awerbuch, S., and Berger, M. (2003). Applying portfolio theory to eu electricity planning and policy making. *IAEA/EET Working Paper No. 03, EET*.

Bar-Lev, D. and Katz, S. (1976). A portfolio approach to fossil fuel procurement in the electric utility industry. *The Journal of Finance*, 31(3):933–947.

Betz, A. (1926). *Wind-energie und ihre ausnutzung durch windmühlen*. Vandenhoeck.

Boettcher, F., Renner, C., Waldl, H.-P., and Peinke, J. (2003). On the statistics of wind gusts. *Boundary-Layer Meteorology*, 108(1):163–173.

Böttcher, F., Peinke, J., Kleinhans, D., Friedrich, R., Lind, P. G., and Haase, M. (2006). Reconstruction of complex dynamical systems affected by strong measurement noise. *Physical Review Letters*, 97:090603.

Brown, R. (1828). Xxvii. a brief account of microscopical observations made in the months of june, july and august 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Philosophical Magazine Series 2*, 4(21):161–173.

Brown, R. (1829). Xxiv. additional remarks on active molecules. *The Philosophical Magazine, or Annals of Chemistry, Mathematics, Astronomy, Natural History, and General Science*, 6(33):161–166.

Burton, T., Sharpe, D., Jenkins, N., and Bossanyi, E. (2001). *Wind energy handbook*. John Wiley & Sons, LTD.

Carapellucci, R. and Giordano, L. (2013). The effect of diurnal profile and seasonal wind regime on sizing grid-connected and off-grid wind power plants. *Applied Energy*, 107:364 – 376.

Cunha, J. and Ferreira, P. (2015). Designing electricity generation portfolios using the mean-variance approach. *International Journal of Sustainable Energy Planning and Management*, 4(0).

Davidson, P. A. (2004). *Turbulence : an introduction for scientists and engineers*. Oxford University Press, Oxford, New York.

Doherty, R., Outhred, H., and O'Malley, M. (2006). Establishing the role that wind generation may have in future generation portfolios. *IEEE Transactions on Power Systems*, 21(3):1415–1422.

Einstein, A. (1905). Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Annalen der Physik*, 322(8):549–560.

Ettoumi, F., Sauvageot, H., and Adane, A. (2003). Statistical bivariate modelling of wind using first-order Markov chain and Weibull distribution. *Renewable Energy*, 28(11):1787 – 1802.

Fawcett, L. and Walshaw, D. (2006). Markov chain models for extreme wind speeds. *Environmetrics*, 17(8):795–809.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, volume 1. Wiley.

Feller, W. (1971). *An introduction to probability theory and its applications. Vol. II.* John Wiley & Sons Inc., New York.

Friedrich, R. and Peinke, J. (1997). Description of a turbulent cascade by a fokker-planck equation. *Physical Review Letters*, 78:863–866.

Friedrich, R., Peinke, J., and Renner, C. (2000a). How to quantify deterministic and random influences on the statistics of the foreign exchange market. *Physical Review Letters*, 84:5224–5227.

Friedrich, R., Peinke, J., Sahimi, M., and Tabar, M. R. R. (2011). Approaching complexity by stochastic methods: From biological systems to turbulence. *Physics Reports*, 506(5):87 – 162.

Friedrich, R., Siegert, S., Peinke, J., Lück, S., Siefert, M., Lindemann, M., Raethjen, J., Deuschl, G., and Pfister, G. (2000b). Extracting model equations from experimental data. *Physics Letters A*, 271(3):217–222.

Fullana, J.-M. and Rossi, M. (2002). Identification methods for nonlinear stochastic systems. *Physical Review E*, 65:031107.

Gardiner, C. (2009). *Stochastic Methods*. Springer-Verlag Berlin Heidelberg.

Gasch, R. and Twele, J. (2012). *Wind power plants*. Springer-Verlag Berlin Heidelberg.

Gottschall, J. and Peinke, J. (2008a). How to improve the estimation of power curves for wind turbines. *Environmental Research Letters*, 3(1).

Gottschall, J. and Peinke, J. (2008b). On the definition and handling of different drift and diffusion estimates. *New Journal of Physics*, 10(8):083034.

Heald, J. P. M. and Stark, J. (2000). Estimation of noise levels for models of chaotic dynamical systems. *Physical Review Letters*, 84:2366–2369.

Hosseini-Firouz, M. (2013). Optimal offering strategy considering the risk management for wind power producers in electricity market. *International Journal of Electrical Power & Energy Systems*, 49:359 – 368.

Huang, Y.-H. and Wu, J.-H. (2008). A portfolio risk analysis on electricity supply planning. *Energy Policy*, 36(2):627 – 641.

Hölling, M., Peinke, J., and Ivanell, S. (2014). *Wind Energy - Impact of Turbulence*. Springer-Verlag Berlin Heidelberg.

Jansen, J., Beurskens, L., and Van Tilburg, X. (2006). Application of portfolio analysis to the dutch generating mix. reference case and two renewables cases, year 2030, se and ge scenario.

Kantz, H., Holstein, D., Ragwitz, M., and Vitanov, N. K. (2004). Markov chain model for turbulent wind speed data. *Physica A*, 342(1–2):315 – 321.

Kemeny, J. G. and Snell, J. L. (1976). *Finite Markov Chains*. Undergraduate texts in mathematics. New York : Springer-Verlag.

Kitzing, L. (2014). Risk implications of renewable support instruments: Comparative analysis of feed-in tariffs and premiums using a mean–variance approach. *Energy*, 64:495 – 505.

Kleinhans, D. (2008). Stochastische modellierung komplexer systeme. Master's thesis, Westfälische Wilhelms-Universität Münster.

Kolmogorov, A. (1941a). The Local Structure of Turbulence in Incompressible Viscous Fluid for Very Large Reynolds' Numbers. *Akademiia Nauk SSSR Doklady*, 30:301–305.

Kolmogorov, A. N. (1941b). Dissipation of Energy in Locally Isotropic Turbulence. *Akademiia Nauk SSSR Doklady*, 32:16.

Krey, B. and Zweifel, P. (2006). Efficient Electricity Portfolios for Switzerland and the United States. SOI - Working Papers 0602, Socioeconomic Institute - University of Zurich.

Kriso, S., Peinke, J., Friedrich, R., and Wagner, P. (2002). Reconstruction of dynamical equations for traffic flow. *Physics Letters A*, 299(2-3):287 – 291.

Kuusela, T. (2004). Stochastic heart-rate model can reveal pathologic cardiac dynamics. *Phys. Rev. E*, 69:031916.

Langevin, P. (1908). Sur la théorie du mouvement brownien. *CR Acad. Sci. Paris*, 146(530-533):530.

Lehle, B. (2011). Analysis of stochastic time series in the presence of strong measurement noise. *Phys. Rev. E*, 83:021113.

Lehle, B. (2013). Stochastic time series with strong, correlated measurement noise: Markov analysis in n dimensions. *Journal of Statistical Physics*, 152(6):1145–1169.

Li, G., Shi, J., and Qu, X. (2011). Modeling methods for genco bidding strategy optimization in the liberalized electricity spot market–a state-of-the-art review. *Energy*, 36(8):4686 – 4700. {PRES} 2010.

Lind, P. G., Haase, M., Böttcher, F., Peinke, J., Kleinhans, D., and Friedrich, R. (2010). Extracting strong measurement noise from stochastic time series: Applications to empirical data. *Phys. Rev. E*, 81:041125.

Losekann, L., Marrero, G. A., Ramos-Real, F. J., and de Almeida, E. L. F. (2013). Efficient power generating portfolio in brazil: Conciliating cost, emissions and risk. *Energy Policy*, 62:301 – 314.

Luhur, M., Peinke, J., Schneemann, J., and Wächter, M. (2014). Stochastic modeling of lift and drag dynamics under turbulent wind inflow conditions. *Wind Energy*, 18:317–337.

Lück, S., Renner, C., Peinke, J., and Friedrich, R. (2006). The markov-einstein coherence length—a new meaning for the taylor length in turbulence. *Physics Letters A*, 359(5):335–338.

Madlener, R. and Wenk, C. (2008). Efficient investment portfolios for the swiss electricity supply sector. *FCN Working Paper No. 2/2008*.

Manwell, J. F., McGowan, J. G., and Rogers, A. L. (2009). *Wind energy explained*. John Wiley & Sons, Ltd.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

Menck, P. J., Heitzig, J., Kurths, J., and Joachim Schellnhuber, H. (2014). How dead ends undermine power grid stability. *Nature Communications*, 5.

Meyer, R. and Christensen, N. (2000). Bayesian reconstruction of chaotic dynamical systems. *Physical Review E*, 62:3535–3542.

Meyer, R. and Christensen, N. (2001). Fast bayesian reconstruction of chaotic dynamical systems via extended kalman filtering. *Physical Review E*, 65:016206.

Milan, P. (2014). The conversion dynamics of wind energy systems treated as a complex stochastic process. Master's thesis, Carl von Ossietzky Universität Oldenburg.

Milan, P., Wächter, M., and Peinke, J. (2013). Turbulent character of wind energy. *Physical Review Letters*, 110:138701.

Nfaoui, H., Essiarab, H., and Sayigh, A. (2004). A stochastic Markov chain model for simulating wind speed time series at Tangiers, Morocco. *Renewable Energy*, 29(8):1407 – 1418.

Papaefthymiou, G. and Klöckl, B. (2008). MCMC for wind power simulation. *Energy Conversion, IEEE Transactions on*, 23(1):234 –240.

Peinke, J., Heinemann, D., and Kühn, M. (2014). *Windenergie- eine turbulente Sache?* Deutsche physikalische Gesellschaft.

Perrin, J. (1913). *Les atomes*. Librairie Félix Alcan.

Pillai, S., Suel, T., and Cha, S. (2005). The perron-frobenius theorem: some of its applications. *Signal Processing Magazine, IEEE*, 22(2):62–75.

Pinson, P., Chevallier, C., and Kariniotakis, G. N. (2007). Trading wind generation from short-term probabilistic forecasts of wind power. *IEEE Transactions on Power Systems*, 22(3):1148–1156.

Platis, A., Limnios, N., and Le Du, M. (1998). Dependability analysis of systems modeled by non-homogeneous Markov chains. *Reliability Engineering and System Safety*, 61(3):235 – 249.

Richardson, L. F. (1922). Weather prediction by numerical process. *Cambridge University press*.

Rinn, P., Lind, P., Wächter, M., and Peinke, J. (2016). Langevin: An r package for stochastic data analysis. *Journal of Open Research Software*. Package and reference manual available at https://cran.r-project.org/web/packages/Langevin/, "Package 'Langevin'", P.Rinn, P.Lind and D.Bastine.

Risken, H. (1984). *The Fokker-Planck Equation*. Springer Berlin / Heidelberg.

Rombauts, Y., Delarue, E., and D'haeseleer, W. (2011). Optimal portfolio-theory-based allocation of wind power: Taking into account cross-border transmission-capacity constraints. *Renewable Energy*, 36(9):2374 – 2387.

Roques, F., Hiroux, C., and Saguan, M. (2010). Optimal wind power deployment in europe—a portfolio approach. *Energy Policy*, 38(7):3245 – 3256. Large-scale wind power in electricity markets with Regular Papers.

Ruszczynski, A. (2006). *Nonlinear Optimization*. Princeton University Press.

Sahin, A. D. and Sen, Z. (2001). First-order Markov chain approach to wind speed modelling. *Journal of Wind Engineering and Industrial Aerodynamics*, 89(3–4):263 – 269.

Shamshad, A., Bawadi, M., Hussin, W. W., Majid, T., and Sanusi, S. (2005). First and second order markov chain models for synthetic generation of wind speed time series. *Energy*, 30(5):693 – 708.

Siefert, M., Kittel, A., Friedrich, R., and Peinke, J. (2003). On a quantitative method to analyze dynamical and measurement noise. *Europhysics Letters*, 61(4):466.

Siegert, S., Friedrich, R., and Peinke, J. (1998). Analysis of data sets of stochastic systems. *Physics Letters A*, 243(5–6):275 – 280.

Smoluchowski, M. (1906). Zur kinetischen theorie der brownschen molekularbewegung und der suspensionen. *Annalen der Physik*, 326(14):756–780.

Suomalainen, K., Silva, C., Ferrão, P., and Connors, S. (2013). Wind power design in isolated energy systems: Impacts of daily wind patterns. *Applied Energy*, 101:533 – 540. Sustainable Development of Energy, Water and Environment Systems.

Sura, P. and Gille, S. T. (2003). Interpreting wind-driven southern ocean variability in a stochastic framework. *Journal of Marine Research*, 61(3):313–334.

Waechter, M., F. Riess, H. Kantz, and J. Peinke (2003). Stochastic analysis of surface roughness. *Europhysics Letters*, 64(5):579–585.

Wen, J., Zheng, Y., and Donghan, F. (2009). A review on reliability assessment for wind power. *Renewable and Sustainable Energy Reviews*, 13(9):2485 – 2494.

Westner, G. and Madlener, R. (2011). Development of cogeneration in germany: A mean-variance portfolio analysis of individual technology's prospects in view of the new regulatory framework. *Energy*, 36(8):5301 – 5313.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83.

Zhu, L. and Fan, Y. (2010). Optimization of china's generating portfolio and policy implications based on portfolio theory. *Energy*, 35(3):1391 – 1402.